Fund Flows and Underlying Returns: The Case of ETFs [*]

Arsenio Staer[†]

This draft: September 16, 2014

Abstract

I investigate the relation between exchange-traded fund (ETF) flows and their underlying securities' returns using a unique database covering U.S. equity ETFs and adjusted for the flow reporting bias. I find a strong positive relation between daily ETF flows and the underlying stocks' returns, suggesting a price pressure effect related to the flow activity. At an aggregate level, vector autoregressive (VAR) tests show that 38% of the flow shock's price change is reversed after five days, which lends support to price pressure hypothesis. These results extend the research concerning the price impact of institutional trades to the novel ETF framework.

JEL Classification: G12, G14, G23
Keywords: ETF, arbitrage, fund, flows, trading volume, VAR, asset basket, asset pricing.

---

[*]I thank Chris Schwarz, Philippe Jorion, Kerry Vandell, Lu Zheng, Zheng Sun, Phil Bromiley, Stergios Skaperdas, Peng-Chia Chiu and participants at the poster session and brown bag presentation at UC Irvine for helpful comments and discussions. All remaining errors are mine.

[†]Correspondence: Department of Finance, California State University at Fullerton, Fullerton, CA 92834. Email: astaer@fullerton.edu

1 Introduction

Like mutual funds, the daily flow activity of exchange-traded funds (ETFs) is very volatile. On September 18, 2007, ETFs tracking the S&P 500 had net inflows of $15.6 billion, which represented 21% of their current market capitalization, while on the previous day, the same S&P 500 ETFs attracted only $2.2 billion. Unlike mutual funds, however, ETF flows are not in the form of money, but rather in the form of the underlying assets. Essentially, the $15.6 billion net inflow on September 18, 2007, meant that specialized market participants[1] bought the equivalent amount of the underlying S&P 500 stocks on the market[2] and exchanged them with the fund for shares of the ETFs. Since there is clearly substantial variability in the magnitude of daily ETF flows, the amount of underlying buying and selling of stocks will vary significantly from day to day. While prior researchers have studied the relation of mutual fund flows with their underlying asset returns (e.g., see Warther 1995, Edelen and Warner 2001, Ben-Rephael et al. 2011), little is known about the relationship between ETF flows and the underlying assets[3]. Given the structural difference between the two industries, it is not clear whether ETF flows will have a similar effect on asset prices.

This relation is extremely important given that exchange-traded funds have enjoyed an accelerated rise to economic prominence since their introduction in 1993, with annual growth rates[4] in excess of 27%. In 2010, all U.S. ETFs had $991 billion in assets under

[1] Authorized Participants (APs) who have signed an agreement with the fund and usually comprise institutional investors and market makers.

[2] If the Authorized Participant already holds the underlying securities, he/she can exchange those securities for the ETF shares without having to buy them on the market. However, according to practitioners (Abner 2011), APs usually transact flow-related orders on the market. Hence, it is reasonable to assume that the majority of flow-related transactions are done in the market and include buying/selling the underlying assets.

[3] To my knowledge the return-flow relation has been studied only once in a paper by Kalaycıoğlu (2004) on the sample of four ETFs. However, the dramatic growth of the ETF market, and hence, size of absolute fund flows and daily trading volume, increase in the number of funds and improvement in data quality warrant a more detailed study.

[4] Computed for the U.S. ETF sample

management while equity ETFs had assets of $476 billion (compared to U.S. equity mutual funds assets of $2,524 billion) . In comparison, in 2006, U.S. ETFs only totaled $422 billion in assets, of which $276 billion was in equity ETFs[5]. Moreover, 14% of institutional investors and 17 of the 20 largest mutual fund complexes report ETFs in their portfolios[6]. The increase in the popularity of ETFs is due to their inherent advantages in index tracking, such as lower expenses, tax efficiency and intraday tradability (Poterba and Shoven 2002), especially when compared with mutual funds (Huang and Guedj 2009).

The key to understanding the difference between mutual fund and ETF flows is the combination of "creation-redemption" activity with "in-kind" transactions. "Creation" occurs when an entity gives the fund the basket of underlying securities and, in exchange, the entity receives an equivalent share of the ETF. "Redemption" is the opposite process. An entity gives the fund a share of the ETF and, in return, receives the equivalent basket of underlying securities. These transactions are known as "in-kind" because the exchange is essentially a barter of equivalent securities[7]. Creation-redemption occurs in the primary market between ETFs and their "Authorized Participants" (APs), usually large broker-dealers and institutions.

Because of the creation-redemption process, APs are required to buy (sell) the underlying assets to create (redeem) ETF shares. Thus, the process itself requires an intervention in the underlying market. The relation of institutional trading and stock prices has been studied extensively in Harris and Gurel (1986) and Lakonishok et al. (1992), Chan and Lakonishok (1993), who find a sizeable but short-lived price impact. While the institutional trading studies report significant results confirming price pressure, the empirical findings documented in the mutual fund literature are mixed. Edwards and

[5] *Investment Company Fact Book* (2009)

[6] "Exchange Traded Funds: Maximizing the Opportunities for Institutional Investors", *Viewpoints*, SSGA, December 2009

[7] This type of transaction is used mostly by ETFs that replicate the index by holding the underlying assets. The in-cash transaction, on the other hand, is mainly employed by ETFs which use derivatives to replicate the index.

Zhang (1998), Warther (2002), Goetzmann and Massa (2003), Rakowski and Wang (2009), Watson and Wickramanayake (2012) use different datasets and methodologies and find no conclusive evidence of price pressure associated with mutual fund flows.

On the other hand, when examining the trades of mutual fund managers, Edelen and Warner (2001) document a statistically significant but weak price pressure using intraday data. Ben-Rephael et al. (2011) employ daily mutual fund flow data from Israel and discover a significant contemporaneous price pressure effect in their sample. In line with this literature, I conjecture that daily net ETF flows cause price pressure on the underlying securities, where a positive flow is related to the positive return in the underlying assets and vice versa.

The positive contemporaneous relation between fund flows and underlying returns cannot necessarily be interpreted as unambiguous evidence for price pressure. There are several hypotheses for the contemporaneous flow and return relation (Edelen and Warner 2001). The first competing explanation is the information hypothesis where positive (negative) information shocks positively (negatively) affect both flows and returns simultaneously. The second possible explanation is based on return chasing investor behavior: Past positive (negative) returns lead investors to invest more (less) in ETFs resulting in positive (negative) flow. Both of these hypotheses provide predictions of a concurrent relation between flows and the underlying asset returns, but no relation between the underlying asset returns and lagged flows. On the other hand, a price pressure explanation implies that the contemporaneous price shock will be followed by a price reversal, which will manifest itself in a negative relation between past flows and returns. Hence, the price pressure hypothesis involves finding that there is a positive relation between the ETF flows and the concurrent underlying asset returns while also finding a negative relation between lagged flows and returns.

To test these predictions, I analyze the relation between ETF flows and their underlying stock returns using a sample of 286 U.S. equity ETFs for the 1993–2010 period. Given that ETFs are specifically constructed to track indices, the association of flows with the underlying stocks, if it exists, should be in proportion to the stocks' weight in the indices. Hence, the underlying index return is an intuitive measure to test the relation between flows and the underlying stock returns. I use data obtained from Bloomberg[8], Center for Research in Security Prices (CRSP) and directly from fund families to construct a database containing index returns and daily shares outstanding. I then adjust the dataset for the $T+1$ reporting bias discussed in Rakowski and Wang (2009), Quinn et al. (2006)[9].

To test the price pressure hypothesis, I perform fund-level and panel regressions with daily index returns as the dependent variable and contemporaneous and lagged ETF flows as the main independent variables. Additionally, given that several funds with similar features can track the same index, I examine commonalities between flows of funds tracking the same index. Next, I study the flow-return relation in a time-series framework using aggregate flows for all U.S. equity ETFs. Finally, in order to extract the transitory price shock from the total contemporaneous price innovation, I employ a vector autoregressive analysis (VAR) on aggregate flows and market returns.

Using individual fund-level regressions and S&P 500 ETF (SPY) as a case study, I find that a 10% increase in flow, expressed as percent change in shares outstanding, corresponds to a 28 basis points (bp) increase in the underlying stocks' returns. The strength of this relation is not monotonic over time. For the 2007–2010 and 2009–2010 subperiods, the relation is 76 bp and 218 bp respectively. Returning to the S&P 500 ETF example at the beginning of this Section and using the coefficient from the 2007–2010 estimation period, the 21% increase in shares outstanding on September 18, 2007

[8]More specifically, *Bloomberg Terminal*

[9]A majority of ETFs report their flows next day after the actual flows occur, however some funds report it the same day. Within the data provided by the *Bloomberg* there is no clear way to identify the exact timing of the flows. Also see Section 3 for discussion.

corresponds to the positive 1.6% S&P500 index return[10,11]. This effect is not limited to the S&P 500 ETF (SPY). On average, the same pattern holds across the 10 largest ETFs. Again, the coefficients are larger in magnitude and more significant in the latter 2007–2010 and 2009–2010 subperiods. Using flows aggregated across all U.S. equity ETFs, I find that a one standard deviation flow shock, which is the equivalent of $2.47 billion of inflows, is associated on average with a 52 basis point shock in the market return (Value-Weighted Index Return[12]) for the 2007–2010 period and with a 54 basis point return shock for the 2009–2010 period. Furthermore, the evidence for price reversal is concentrated in the top 10 ETFs . In panel regressions for the top 10 ETFs during 2007–2010, the relation between lagged flows and returns is significantly negative, with a one standard deviation change in flows corresponding to a five basis points price reversal, which represents 22% of the contemporaneous coefficient's magnitude.

Employing a VAR analysis on aggregate flows and market returns for the 2007–2010 period, I also find a significant price reversal effect. Cumulative impulse response functions (CIRFs) show that after five days, there is a price reversal of 38% of the initial shock with the remaining 62% representing a permanent price change. Economically, a one standard deviation shock in aggregate ETF flow ($\sigma_{flow} = 0.015$) is related to temporary and permanent shocks of 27 bp and 44 bp in market returns respectively.

My findings contribute to the literature in a number of ways. First, this paper explores a novel proxy for institutional trading based on the particular features of ETFs and is the first study to combine daily data for a large cross-section of funds across over a 17-year period. Second, it adds to the literature on the effect of institutional trading on prices, including providing evidence for the "price pressure" hypothesis (Edelen and Warner

[10]Price pressure applies to the underlying assets' prices which are almost immediately reflected in the index. Hence, index return provides a convenient statistic that aggregates price impact for each individual stock. Further discussion is in Section 3

[11]The actual S&P500 index return on that date was 2.9%

[12]Obtained from Center of Research in Stock Prices (CRSP)

2001, Goetzmann and Massa 2003). Third, the results of this paper are also related to the microstructure literature, particularly to price deviations from fundamentals due to trading (Biais et al. 2005). Exchange-traded funds could be considered as derivatives or contingent claims on other securities, in this case, the underlying stocks comprising the index (Roll et al. 2010). Consequently, the results add a new angle to the long-standing debate about whether derivative trading has an effect on the prices of the underlying assets (Subrahmanyam 1991, Roll et al. 2012).

The remainder of the paper proceeds in the following way. Section 2 outlines the relevant literature. Next, I describe the characteristics of the data in Section 3. Section 4 outlines the empirical analysis of the flow-return relation. Section 5 concludes. Appendix A contains auxiliary tables.

2 Related Literature

In the classical finance, demand curves are elastic and stocks are near perfect substitutes for each other. Hence, theoretically, the sales and purchases of large amounts of stocks should not move prices. However, since 1970s, there exists mounting empirical evidence (Scholes 1972) that demand curves are downward sloping and prices fluctuate due to demand shocks. To study this relationship, researchers have generally used two approaches.

In the first approach, Harris and Gurel (1986), Shleifer (1986) examine the effect of demand shocks on stock returns using index inclusions and exclusions[13]. The authors conclude that the index inclusions and exclusions, which presumably[14] are information-free demand shocks, are associated with the underlying stock's positive and negative abnormal

[13]For a look at asymmetry of index inclusion and exclusion effects, see Chen et al. (2004). For the examination of option trading impact on abnormal return associated with index inclusion, see Chen et al. (2013)

[14]For discussion on information content of index inclusions, see Denis et al. (2003)

returns respectively. More recently, Petajisto (2009) introduces financial intermediaries in the demand model to explain the empirically observed index inclusion effects.

The second approach, which is used in this paper, explores the effect of demand shocks on prices using market variables such as excess order flow (Boehmer and Wu 2008), block trades (Lakonishok et al. 1992, Chan and Lakonishok 1993) and institutional flows (Wermers 1999, Cai and Zheng 2004). Within this second approach, the relation between investor flows and returns on the aggregate level is studied by Boyer and Zheng (2009) who find a significant positive association between quarterly cash flows for some investor groups and market returns but discover no conclusive evidence of price pressure. Warther (1995), Goetzmann and Massa (2003) focus on mutual fund flows and returns using different samples and methodologies and also do not find significant evidence of price pressure. Edwards and Zhang (1998) examine flows into equity and bond mutual funds and their relation with market returns and find no evidence of price pressure except during the 1971–1981 period when massive redemptions negatively affected market returns. Rakowski and Wang (2009) use a large mutual fund database of fund-specific daily flows from Lipper to document the flow-return relation for the period from 2000 to 2006 and also fail to find support for the price pressure hypothesis. Furthermore, applying individual fund regressions on weekly and monthly horizons to ETFs, Kalaycıoğlu (2004) examines a sample of four ETFs and does not find a significant relation between flows and returns on the fund level.

On the other hand, Edelen and Warner (2001) document a significant relation between lagged flows and intraday returns indicating evidence for price pressure. In a more recent paper by Ben-Rephael et al. (2011), the researchers investigate the association between aggregate Israeli mutual fund flows and underlying market returns using daily data. The authors find a significant correlation between concurrent market returns and mutual fund flows and a negative relation between lagged fund flows and returns. Their results are consistent with a price reversal and, hence, with the price pressure hypothesis. My paper contributes to the debate over price pressure in investment fund flows by extending it to the ETF context.

Although my paper is closely related to the mutual fund flow and return literature, flows in the ETF and mutual fund industries are structured differently. In the mutual fund case, managers respond to daily creation and redemption requests, either by investing the money in the fund's portfolio in case of creation or by selling portfolio shares in the market and returning proceeds to the investor in case of redemption. In the ETF case, daily creation and redemption is transacted with underlying stocks instead of cash. These "in-kind" transactions occur in the primary market between the fund issuer and Authorized Participants (APs), who are usually large institutional investors and market makers who have signed a special agreement with the fund. APs then operate in the stock market to buy or sell the transacted stocks underlying the ETF shares. In essence, in the ETF case, flow-related trading activity is shifted towards APs and not the fund itself. The only difference with the mutual fund case is that the flow-related trades are executed not by the fund manager but by the APs.

Furthermore, ETFs can be considered securities representing weighted portfolios of other assets or asset baskets (Gorton and Pennacchi 1993), which have been studied extensively. For example, Subrahmanyam (1991) examines uninformed investor trading strategies (Kyle 1985) when facing an informed investor and concludes that asset baskets mitigate adverse selection costs. Trading in the S&P 500 index futures, which started in 1982 and quickly achieved considerable volume and investor attention (Schwert 1990), presents empirical evidence that investors assign considerable value to asset baskets even though their payoffs are redundant in classical finance. Moreover, asset baskets are also contingent claims[15] on other securities' payoffs, and, hence, can be regarded as derivatives. Roll et al. (2012) investigate the macroeconomic forecasting power of trading volume of

[15]The contingent claims include: options, futures, forwards and other financial derivatives. ETFs can be considered contingent claims too as their prices are dependent on the weighted payoffs of their holdings.

S&P 500 index-linked derivatives and include asset baskets such as ETFs. Furthermore, Pearson et al. (2007) study the effect of option trading on the underlying stock return variability. My findings contribute to this literature by showing that flows in the derivative market (ETFs) are associated with price pressure in the underlying asset markets.

Another strand of literature focuses on pricing characteristics of ETFs such as fund premium and tracking error. Huang and Guedj (2009), Doran et al. (2006) investigate differences in performance and tracking error between index mutual funds and ETFs, while Poterba and Shoven (2002) analyze tax aspects of ETFs. Furthermore, fund premiums for international vs. domestic ETFs is studied by Delcoure and Zhong (2007), Milonas and Rompotis (2010) with an extension into the arbitrage and liquidity context by Ackert and Tian (2008). Additionally, the advent of leveraged ETFs has sparked research into tracking error in leveraged funds, with papers by Lu et al. (2009), Charupat and Miu (2011). Furthermore, Tse et al. (2006), Hseu et al. (2007), Schlusche (2009) explore price discovery between ETFs and other index-linked securities and find that price innovation is usually faster in the ETF and futures markets than in the underlying indices.

Given that ETFs share many characteristics with more mature closed-end funds but track the underlying index much closer (Shin and Soydemir 2010), researchers have embraced them as a valid way to study the interaction between asset baskets and their components. Adverse selection and liquidity in ETFs are studied by Chen and Strother (2008), who investigate price discovery in Asian markets, by Hamm (2010), who analyze the link between ETFs and component liquidity, and by Richie and Madura (2007) with their exploration of ETF creation and the underlying stock liquidity. The general agreement is that the effect of ETFs on the pricing characteristics of the constituent stocks is significant both economically and statistically.

3 Data and Summary Statistics

Following Kalaycıoğlu (2004) and Svetina and Wahal (2008), I use *Bloomberg Terminal* as the source of my ETF data for the period starting in January 1993[16] and ending in July 2010. At the end of the sample period, the global ETF universe consisted of 5,126 exchange traded funds[17]. Among them, 891 ETFs trade on the U.S. markets. The data on ETFs includes prices, NAV[18], shares outstanding, underlying index tickers and returns and asset-based style descriptions. Available data only covers funds that have survived till the end of the sample, hence the database excludes "dead" funds. However, this fact is unlikely to introduce a survivorship bias, as low assets under management (AUM) and generally low flows are cited most often as the cause of ETF delisting[19]. Furthermore, I do not investigate hypotheses pertaining to fund characteristics but instead center on the empirical regularities associated with the funds' activity.

The shares outstanding data for ETFs[20] are also present in Center of Research in Stock Prices (CRSP), but the update frequency is usually monthly instead of daily with a few exceptions. I cross-check the data on shares outstanding with the CRSP stock database and fill in the missing data where possible[21]. I also parse ETF websites and verify the data by hand, especially for fund families following "T" accounting such as ProShares[22]. The shares outstanding and NAV data are adjusted for splits, reverse splits and distributions.

The index returns are a perfect instrument to measure price impact on the underlying assets. Furthermore, index returns are calculated in real-time, hence any price impact

[16] First ETF, SPDR S&P 500 ETF (SPY), was introduced on January 22, 1993

[17] Some European ETFs are double counted due to trading under different stock tickers in different markets.

[18] Net Asset Value or the amount per share of the total value of the fund's securities less any liabilities.

[19] Usually, mutual fund flow researchers, for instance Edelen and Warner (2001), Rakowski and Wang (2009), do not consider survivorship bias germane to the price pressure hypothesis tests. Furthermore, they usually delete funds with very low TNAs due to erratic flows

[20] The ETFs are assigned a share code (*shrcd*) 73 in *CRSP*

[21] For example, 2004–2008 data for NASDAQ100 ETF (QQQ) is absent from *Bloomberg* and present in CRSP as of August 2010.

[22] Interview on August 18, 2010 with ProShares representative.

in the underlying asset basket is transferred to the index level almost immediately (Tse et al. 2006). As of July 2010, U.S. equity ETFs tracked 275 different indices. I obtain daily adjusted prices for these indices from *Bloomberg* and cross-check them with other publicly available sources to ensure data integrity.

In this paper, I focus on U.S. equity ETFs due to the transparency of the creation-redemption process, overall importance for the economy and the amount of academic research. The ETF industry in Europe and Asia is also well-developed, but the regulatory requirements are less transparent and the requirement to hold physical assets is not as strict as in the U.S. Out of total 5,126 ETFs at the end of the sample period, 923 are traded in the U.S. markets, and out of them, 349 are U.S. equity ETFs.

ETFs can also employ leverage via swaps or futures[23], eschewing the "in-kind" creation-redemption process for the cash-for-shares exchange. Usually, leverage is used to provide long or short multiples of the underlying index returns and is explicitly stated in the ETF name and prospectus. I screen my ETF sample names and descriptions for these keywords: "Bear", "Ultra", "Short", "Inverse", "2x", "3x", "-2x". After screening, 63 leveraged funds are removed and 286 non-leveraged U.S. equity ETFs remain in the sample. The final sample represents about 40% of the total net assets and 32% of the total number of funds in the U.S. ETF universe.

In Figure 1, I plot the number of funds, underlying indices and total net assets for U.S. equity ETFs over time. The recent decrease in the growth rate of the population of funds and tracked indices suggests that the U.S. equity ETF industry has reached a relatively stable amount of index coverage. Specifically, the number of funds has started to outpace the amount of indices tracked (Svetina and Wahal 2008). This suggests an increased competition between funds and a scarcity of available indices to track.

Given the rapid growth and relatively young age of the ETF industry, empirical phenomena related to its flow activity may also vary considerably across time. Partitioning

[23] "ProShares ETFs: Frequently Asked Questions about Geared Funds", *ProShares.com*, April 13, 2010.

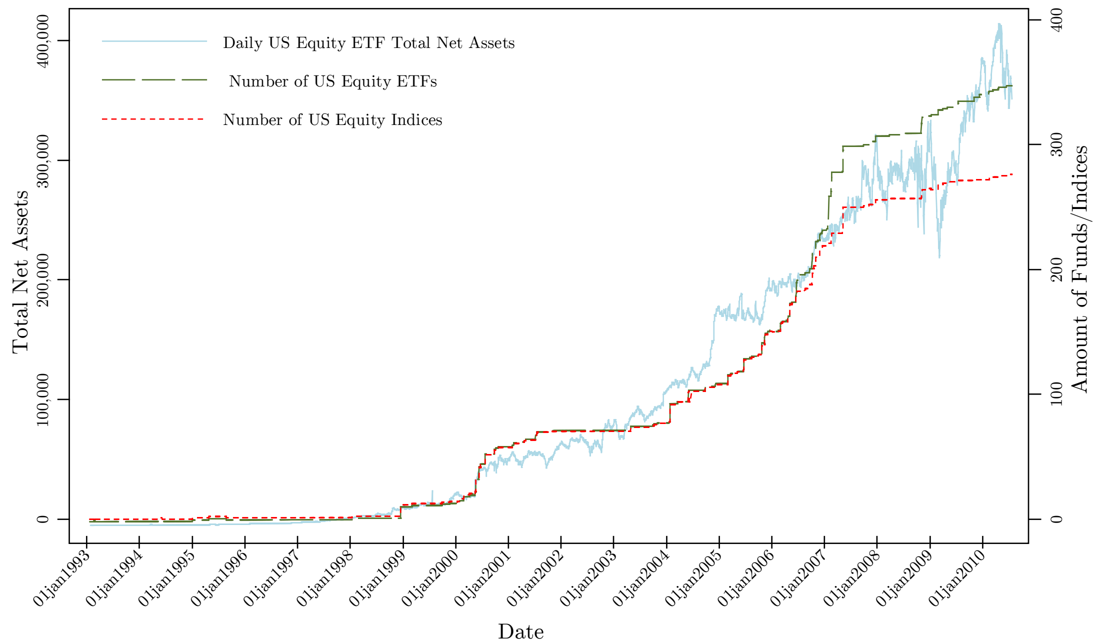


Figure 1. US Equity ETF Total Net Assets, Number of Funds and Indices
This figure displays time series for total net assets, number of funds and underlying indices for U.S. Equity ETFs during the period from 1993 to 2010. Total net assets are in $ Million. Given that several funds may track the same index, the total number of funds can be larger than the total number of indices. Only indices that were tracked by at least 1 fund at that particular date are included.

the sample period into subperiods is helpful in determining whether the effects are stable across time when testing the hypothesis of interest. The period of 2007–2010 is the first period when the upward trend in total net assets becomes less prominent and flows become more stationary. On the other hand, this period also contains several financial crises and a recession which might affect the flow-return relation. Consequently, I also include a shorter and more recent 2009–2010 subperiod which offers more market stability. My choice of subperiod length is comparable to mutual fund studies, such as Goetzmann and Massa (2003), who use two years of daily data in their analysis.

To analyze the total net asset distribution of my sample, I take the last month of the sample, average total net assets for each ETF and sort the resulting averages into deciles. As documented in Table 1, the total net asset distribution exhibits a significant positive skewness. More specifically, the TNA for the 10^{th} decile totals $251 billion, which corresponds to 72.9% of my sample's TNA, while the TNA for 1^{st} decile is only $0.4 billion,

Table 1. Sample Summary Statistics

This table contains summary statistics for the U.S. equity ETFs sample ranked by deciles as of July 2010. Funds are ranked from one to ten by the average total net assets (TNA) during July 2010. Minimum, maximum, average and total TNA are in $ millions and specify the distribution parameters of the total net assets in each decile. % of Total Sample TNA is the ratio of the decile TNA to the total sample TNA.

TNA Decile	No of funds	Minimum TNA	Maximum TNA	Average TNA	Total TNA	% of Total Sample TNA
1	29	3	22	14	415	0.1%
2	28	23	37	31	879	0.3%
3	29	38	54	46	1,345	0.4%
4	28	54	99	71	2,000	0.6%
5	29	100	150	125	3,618	1.1%
6	28	152	225	181	5,061	1.5%
7	29	227	405	293	8,508	2.5%
8	28	407	1,055	602	16,858	4.9%
9	29	1,086	2,760	1,876	54,405	15.8%
10	28	2,781	68,415	8,957	250,794	72.9%
Total	285	3	68,415	1,207	343,882	100%

which amounts to 0.1% of the sample. Furthermore, the maximum and minimum TNA for the 10th decile exhibit a much wider range compared to the other deciles.

Given that ETF flows depend on investors' interest in the fund, which in turn is correlated with the fund size, it is reasonable to expect higher absolute flows for larger ETFs. Moreover due to the skewness of the TNA distribution, an analysis of a smaller subset of the high TNA ETFs should provide some inferences while maintaining representativeness. In Table 2, I show summary statistics for the top 10 ETFs ranked by TNA in July, 2010. The largest and oldest fund in the sample, SPDR S&P 500 ETF Trust, has nearly 3 times the TNA of the 2nd largest fund, iShares S&P 500 Index Fund. All of the top 10 ETFs, with the exception of the Vanguard ETF (VTI), were formed on or before 2000. Similar to Table 1, top 10 ETFs hold 51% of the total net assets for the equity ETFs. The data on the largest ETF funds for each Investment Advisor is reported in Appendix A-1.

Importantly, exchange-traded funds often report using "T+1" accounting[24], which is widely followed in the mutual fund industry as described in Quinn et al. (2006). This accounting or dating methodology implies that the shares outstanding reported to the data

[24] Interview with Mr. Don Suskind (PIMCO), and phone interviews with other ETF families

Table 2. Top 10 US Equity ETFs

Descriptive statistics for top 10 equity ETFs traded in U.S. market for July, 2010. Inception is the date when the fund started trading. Average TNA is the average over last 5 trading days of the sample in $ Millions. Style is based on market capitalization of the tracked index components and is obtained from *Bloomberg Terminal*. The proportion of Fund TNA relative to the total equity ETF TNA from Table 1 is denoted by % of Total TNA.

Fund Name	Ticker	Inception	Style	Index Ticker	Average TNA	% of Total TNA
SPDR S&P 500 ETF Trust	SPY	21-Jan-93	Large	SPX	67,249	19.6%
iShares S&P 500 Index Fund	IVV	15-May-00	Large	SPX	21,672	6.3%
PowerShares QQQ NASDAQ 100	QQQ	4-Mar-99	Large	NDX	17,989	5.2%
Vanguard Total Stock Mkt	VTI	14-Apr-05	Large	MSCIBM	13,838	4.0%
iShares Russell 2000	IWM	22-May-00	Small	RTY	12,616	3.7%
iShares Russell 1000 Growth	IWF	22-May-00	Large	RLG	10,214	3.0%
iShares Russell 1000 Value	IWD	22-May-00	Large	RLV	8,429	2.5%
SPDR S&P Midcap 400 ETF	MDY	26-Apr-95	Mid	MID	8,313	2.4%
SPDR DJIA Trust	DIA	13-Jan-98	Large	INDU	7,698	2.2%
iShares S&P Midcap 400	IJH	22-May-00	Mid	MID	7,086	2.1%
Total					175,106	50.9%

vendor can have a one day lag. In other words, if today's change in shares outstanding is an inflow of 2 million shares, this inflow will not be registered until the next day. Conversely, if the dating is current, the inflow of 2 million shares is registered today. Within the data provided by the platform there is no clear way to identify the lag[25].

There are several ways to identify the dating methodology used by the fund. The first approach is used by Edelen and Warner (2001) who compare N-30b-2 semiannual filings for mutual funds. This approach can also be applied to certain types of ETFs like ETF trusts. The second approach analyzes N-CSRS semiannual filings which are also applicable to open-end ETFs. Shares outstanding reported in the filings have to match the correct date due to legal requirements. To complicate matters further, funds can change their date reporting without making the change public. However, these changes do not seem to occur more than one or two times a year.

For mutual fund industry, this distinction is not that important because the data on daily flows is uncommon and, on monthly level, this bias is not highly significant (Quinn

[25] In fact, *Bloomberg* was not aware of this issue when contacted by the author in May 2010

et al. 2006). For ETFs where shares outstanding have to be posted daily by regulatory requirements, this distinction is crucial. Failure to account for this disparity introduces a bias in the studies that use daily ETF shares outstanding. I analyze N-CSRS and N-30b-2 filings for each fund family and match the dates for shares outstanding in the filings to the dates in the sample.

4 Empirical Analysis

Given the complexity of exchange-traded fund creation-redemption process, it is important to describe the flow mechanism in detail. There are two main causes for ETF flow (Abner 2011, Agapova 2010). In the first case, institutional investors, who want to buy ETF shares and want to avoid exerting price impact in the ETF market, submit an order to buy ETF shares with the Authorized Participant (AP) who then buys the underlying assets on the market. At the end of the day, AP brings purchased underlying assets and exchanges them for the ETF shares directly with the fund. In the case of selling ETF shares, the process is the reverse — investors transfer the ETF shares to the AP who then exchanges them with the fund for the underlying assets, sells the assets on the market and returns the proceeds to the investor. Usually, however, APs, after receiving the sell order for ETF shares, sell short the underlying and return the proceeds to the investors. Later the same day, the APs then exchange the ETF shares for the underlying assets with the funds and cover the short sale.

In the second case, an excess of demand (or supply) of ETF shares in the market leads to an increasing premium (or discount), which generates arbitrage opportunities for Authorized Participants. The APs then create (or redeem) ETF shares using the underlying asset basket to exploit the arbitrage opportunity. Given that APs are usually market makers, broker/dealers or large institutions with access to low transaction cost trading, they can relatively easily buy (or sell) the underlying assets basket to create (or redeem) the ETF shares. This process occurs daily and APs stand ready to take advantage of an arbitrageable spread between the ETF price and the NAV. In the case of a premium[26], once the spread exceeds transaction costs, orders are executed to buy the underlying basket and to sell short the ETF shares, while a creation order is sent to the fund. At 4:00 p.m. EST, the fund calculates the NAV and executes the creation order by receiving the basket of the underlying securities and sending newly created shares to the APs. Next, APs cover the short sale and receive the spread minus the transaction costs. The case for the discount is the reverse.

Given that institutional order flow can move prices (Lakonishok et al. 1992, Shleifer 1986), a large inflow (outflow) in an ETF should result in an upward (downward) pressure on the underlying stock prices. If the effect is transitory (Froot et al. 2001), then the underlying stocks' prices will revert to their pre-shock levels. On the other hand, if the effect is permanent, then the prices will not revert. Price reversal can be identified by a negative relation between lagged flows and contemporaneous returns. If the price change is due to flow pressure, the magnitude of the relation between flows and returns is high on the contemporaneous basis and equal or lower on the lagged basis. Summarizing,

Hypothesis 1. Price Pressure. *ETF flows are positively and significantly related to the contemporaneous underlying stock returns, while lagged ETF flows are negatively and significantly related to the contemporaneous underlying stock returns.*

Mutual fund researchers use two main approaches to study similar hypotheses. The first approach is based on analyzing flow-return relation for each fund or groups of funds. For example, Rakowski and Wang (2009) implement VAR regressions for each fund, while Warther (1995) group mutual fund flows into subcategories and analyze the relation between flows for each category and their underlying returns. The second approach is based on using aggregate flows for the mutual fund industry on different frequencies,

[26] ETF premium is defined as the positive difference between ETF price and net asset value (NAV). ETF discount is the reverse.

from quarterly (Boyer and Zheng 2009, Jank 2012) to daily (Edelen and Warner 2001). Aggregate flow studies usually cannot isolate the flow-return relation within a specific subset of the stock market. For instance, a negative flow correlation between different similarly sized funds will result in a low or zero net aggregate flow (Warther 1995, Cao et al. 2008), but fund-level flows might still be related to their specific underlying stock returns. Hence, analysis of the cross-sectional behavior of ETF flows might provide additional evidence for the price pressure hypothesis.

To test Hypothesis 1, I combine the two aforementioned approaches. First, I run individual and panel regressions on the cross-section of ETFs using a methodology similar to the one used by Ben-Rephael et al. (2011), Goetzmann and Massa (2003). I also analyze commonalities in the flows of funds tracking the same index. Finally, I aggregate flows across ETFs and analyze them using time series and vector autoregressive regressions. In summary, I employ several levels of flow aggregation: individual fund, index or all funds tracking the same index, and aggregate flow for the whole sample.

4.1 Variable Construction

In the mutual fund literature, there are several ways to measure fund flows. For instance, Edelen and Warner (2001) use the percentage change in assets under management (AUM) less the one-day percentage change in NAV, while Goetzmann and Massa (2003), Oh and Parwada (2007) use changes in assets under management scaled by a rolling market capitalization. In the case of ETFs, I use slightly different measures, given access to the data on shares outstanding. I introduce them and summarize their characteristics below.

The flow variable is based on the change in daily shares outstanding scaled by the previous day's total shares outstanding,

$$Flow_t = \frac{Shrout_t - Shrout_{t-1}}{Shrout_{t-1}} \tag{1}$$

where $Shrout_t$ represents total ETF shares outstanding at t. I will refer to $Flow_t$ as fund flow or flow. This formulation is independent of the index price, NAV or fund price and allows to compute flows via changes in shares outstanding instead of changes in assets under management. The disadvantage is that the economic magnitude of the flow is not observable and this measure has to be normalised by NAV or total net assets when summed or averaged across funds.

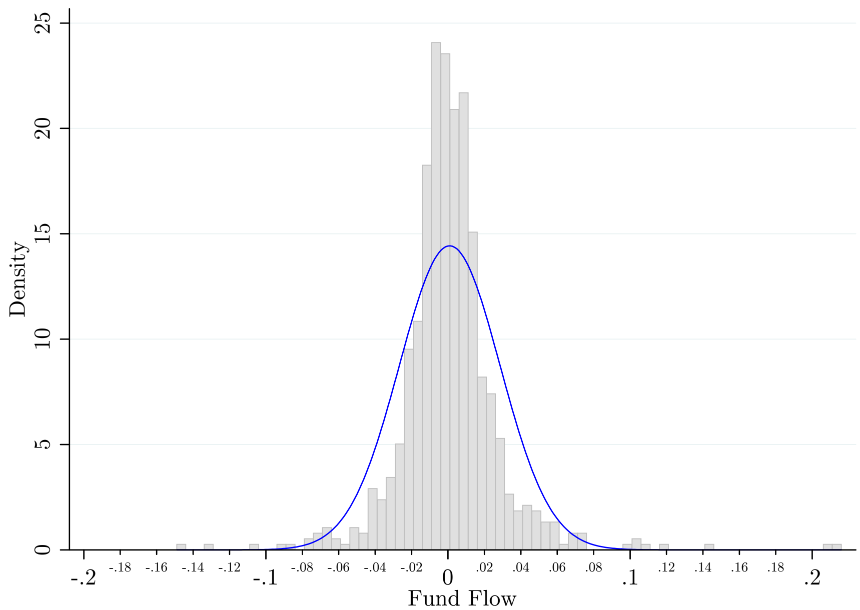


Figure 2. ETF Flow Distribution

Flow distribution corresponds to S&P 500 ETF for the 2007-2010 period. Fund flows are specified as in equation (1). Normal distribution is superimposed on the histogram.

Figure 2 shows a distinctly non-normal distribution of flows for the S&P500 ETF (SPY) over the 2007–2010 period. The flow distribution is similar across all funds; however, funds with lower assets usually exhibit lower frequency of creation-redemption activity, and hence higher kurtosis. For instance, the S&P 500 ETF (SPY) has 2 days with no flows in 2009, while the next largest fund in the sample, iShares S&P 500 ETF (IVV), has 72 days with no flows in the same year.

Given structural differences between ETFs, mutual funds and closed-end funds, it is important to analyze the relations between flows and pricing variables to avoid spurious associations. Table 3 presents correlations between contemporaneous and lagged flows and different index and fund pricing variables using the S&P 500 ETF daily data for 2007–2010[27]. Flow shows a strong correlation with the change in shares outstanding ($\rho_{Flow,\Delta Shrout} = 0.98$) and exhibits a small but significant positive correlation with the index return ($\rho_{Flow,\ Index\ Ret} = 0.11$) which supports Hypothesis 1. Flows also exhibit a weakly significant and positive autocorrelation ($\rho_{Flow_t,Flow_{t-1}} = 0.08$).

One of the often cited reasons for the creation-redemption activity in ETFs is the arbitrage of the difference between fund price and the NAV, or the spread, as shown in the following equation,

$$Spread_t = \frac{P_t^{etf} - NAV_t}{NAV_t} \tag{2}$$

One could argue that an increase in the index level is associated with the increase in the NAV ($\rho_{Index\ Price,\ NAV} = 0.99$) and is correlated with the negative spread ($\rho_{Index\ Ret,Spread_t} = -0.197$) given a constant ETF price. Next, the increase in the absolute spread induces APs to exploit this arbitrage opportunity (if the absolute spread exceeds transaction costs) resulting in redemption of ETF shares and fund outflows. Following this argument, $Spread_t$ should exhibit significant negative correlation with the flow. However, Table 3 shows that neither the contemporaneous or lagged flows are either statistically or economically correlated with the spread ($\rho_{Flow,Spread} = 0.008$ and $\rho_{Flow_{t-1},Spread_t} = 0.052$

[27]Correlations for the whole sample period of 1993–2010 are similar but smaller in absolute value. Creation-redemption facility was used in less than half of trading days every year from 1993 to 2000, hence the flow correlations in the earlier periods of the sample are necessarily biased downward.

Table 3. Flow Correlations

The table shows correlations between variables for the S&P500 ETF during years 2007–2010. The $Flow_t$ and $Flow_{t-1}$ are the contemporaneous and lagged daily percent change in shares outstanding from equation (1). NAV is Net Asset Value reported by the fund by the end of the day. ETF Price is the ETF share price at the close. Shares outstanding contemporaneous and lagged are $Shrout_t$ and $Shrout_{t-1}$ respectively. The daily change in shares outstanding is represented by $\Delta Shrout_t$. The difference between ETF price and NAV normalized by the NAV at time t is denoted by $Spread_t$ as in equation (2). The p-values are in parenthesis below the correlation parameter estimates.

Variables	$Flow_t$	$Flow_{t-1}$	NAV	ETF Price	$Shrout_t$	$\Delta Shrout_t$	Index Price	Index Ret	$Spread_t$
$Flow_t$	1								
$Flow_{t-1}$	0.076	1.000							
	(0.035)								
NAV	0.057	0.048	1.000						
	(0.115)	(0.180)							
ETF Price	0.057	0.049	0.999	1.000					
	(0.115)	(0.175)	(<0.001)						
$Shrout_t$	0.044	0.054	-0.794	-0.793	1.000				
	(0.219)	(0.131)	(<0.001)	(<0.000)					
$\Delta Shrout_t$	0.975	0.053	0.052	0.052	0.052	1.000			
	(<0.001)	(0.138)	(0.147)	(0.147)	(0.149)				
Index Price	0.056	0.046	0.999	0.999	-0.792	0.051	1.000		
	(0.123)	(0.197)	(<0.000)	(<0.000)	(<0.001)	(0.156)			
Index Ret	0.108	-0.062	0.023	0.021	-0.014	0.120	0.023	1.000	
	(0.002)	(0.085)	(0.519)	(0.558)	(0.695)	(0.001)	(0.524)		
$Spread_t$	0.008	0.052	0.042	0.052	-0.034	0.005	0.041	-0.197	(1.000)
	(0.818)	(0.148)	(0.252)	(0.152)	(0.355)	(0.890)	(0.259)	(<0.000)	

respectively). This provides evidence against the argument that fund flows are related to the index return through the fund spread.

A different way to measure flow is to compute dollar flows as seen in Ben-Rephael et al. (2011). Formally, dollar flow is a change in the shares outstanding multiplied by the NAV calculated at 4 p.m. EST on the day the flow occurs,

$$DFlow_t = (Shrout_t - Shrout_{t-1})NAV_t \tag{3}$$

where $DFlow$ is the dollar flow, $Shrout$ is shares outstanding at time t and NAV_t is the fund net asset value at time t. This is the exact monetary equivalent of the shares that the APs exchanged with the fund.

Table 4. Dollar Flow Correlations

The table shows correlations between variables for the S&P500 ETF during years 2007–2010. The $DFlow_t$ and $DFlow_{t-1}$ are the contemporaneous and lagged daily dollar flows from equation (3). NAV is Net Asset Value reported by the fund by the end of the day. ETF Price is the ETF share price at the close. Shares outstanding contemporaneous and lagged are $Shrout_t$ and $Shrout_{t-1}$ respectively. The daily change in shares outstanding is represented by $\Delta Shrout_t$. The p-values are in parenthesis below the correlation parameter estimates.

Variables	$DFlow_t$	NAV	ETF Price	$Shrout_t$	$\Delta Shrout_t$	Index Price	Index Ret	$Flow_t$	ETF Ret	$Spread_t$
$DFlow_t$	1.000									
NAV	0.046	1.000								
	(0.205)									
ETF Price	0.046	1.000	1.000							
	(0.204)	(<0.000)								
$Shrout_t$	0.053	-0.794	-0.793	1.000						
	(0.147)	(<0.000)	(<0.000)							
$\Delta Shrout_t$	0.972	0.053	0.053	0.053	1.000					
	(<0.000)	(0.149)	(0.148)	(0.148)						
Index Price	0.045	1.000	1.000	-0.792	0.051	1.000				
	(0.217)	(<0.000)	(<0.000)	(<0.000)	(0.158)					
Index Ret	0.100	0.023	0.021	-0.014	0.121	0.023	1.000			
	(0.006)	(0.519)	(0.559)	(0.696)	(0.001)	(0.524)				
$Flow_t$	0.990	0.057	0.057	0.045	0.976	0.056	0.108	1.000		
	(<0.000)	(0.115)	(0.115)	(0.219)	(<0.000)	(0.123)	(0.002)			
ETF Ret	0.105	0.020	0.019	-0.012	0.120	0.020	0.987	0.113	1.000	
	(0.004)	(0.584)	(0.603)	(0.727)	(0.001)	(0.590)	(<0.000)	(0.002)		
$Spread_t$	0.015	0.042	0.052	-0.033	0.014	0.041	-0.197	0.008	-0.088	1.000
	(0.679)	(0.252)	(0.152)	(0.355)	(0.693)	(0.259)	(<0.000)	(0.819)	(0.015)	

Table 4 shows the correlations between dollar flows and other pricing variables for the S&P 500 ETF (SPY) for the 2007–2010 period. Dollar flows are computed using NAV, hence they are implicitly linked to the index values, but the correlation is statistically insignificant with the p-value of 0.206. Moreover, dollar flows exhibit high absolute correlations with the change in shares outstanding, flows as measured in equation (1), index returns, and, unsurprisingly, ETF returns. Both $DFlow$ and $Flow$ are highly correlated ($\rho_{Flow,DFlow} = 0.99$). The correlation between dollar flows and index returns is high compared to the other index return correlations ($\rho_{DFlow,\ Index\ Ret} = 0.1$) with p-value of 0.006, which is close in magnitude to the flow-return correlation ($\rho_{Flow,\ Index\ Ret} = 0.11$). The association of dollar flows with the spread is similar to that of the flows and is not significant with p-value of 0.82. It would be safe to say that the dollar flows, like the flow from equation (1), do not exhibit correlations with the index returns through spreads.

Figure 3 shows the time series for daily shares outstanding and NAV for the S&P 500 ETF (SPY) for the 2007–2010 period. There is no clear evidence of a positive correlation between shares outstanding and NAV, and it is clear that a long-run trend in NAV does not necessarily correspond to the same trend in shares outstanding. In fact, the correlation from the Table 4 is significantly negative ($\rho_{Shrout,\ NAV} = -0.794$). Interestingly, however, first differencing the shares outstanding variable reduces the absolute correlation between NAV and shares outstanding to almost to zero ($\rho_{\Delta Shrout,\ NAV} = 0.053$).

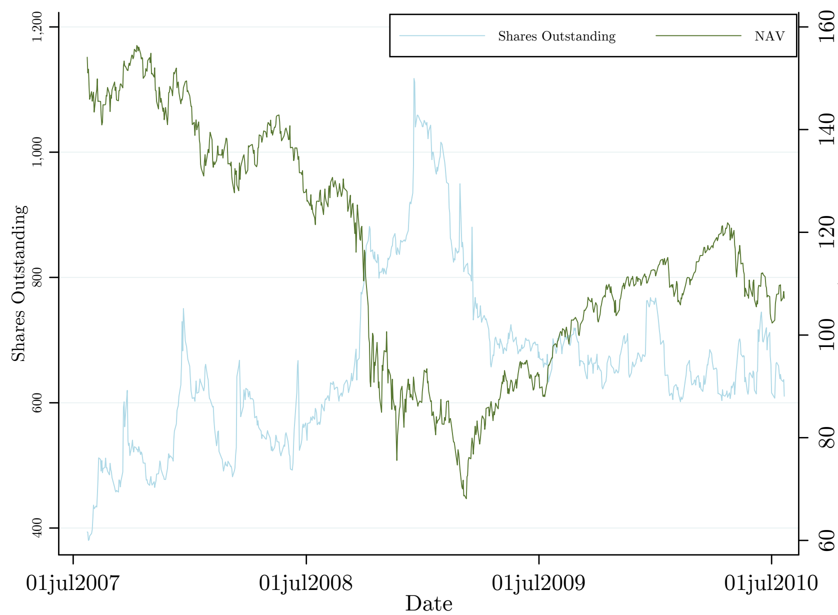


Figure 3. Shares Outstanding and NAV
Time series represent shares outstanding and NAV for S&P 500 ETF for the 2007-2010 period. Shares outstanding are in millions of shares. NAV is in dollars per share.

4.2 Individual Regressions

To test Hypothesis 1, I perform individual fund regressions of daily index returns on ETF flows. I follow the approach used by Edelen and Warner (2001), Ben-Rephael et al. (2011), which consists in selecting five daily lags for returns and five daily lags for the flow variable. Due to the serial autocorrelation of daily index returns (Ackert and Tian 2008), I estimate the regressions adjusting for heteroskedasticity and autocorrelation with standard error correction for five lags implemented by Newey and West (1987). I conduct the preliminary analysis using two ETFs from Table 2: S&P 500 ETF (SPY) and NASDAQ-100 ETF (QQQ). Given the high concentration of ETF assets in the top decile, individual regressions for these ETFs with large total net assets should still be economically meaningful. Next, I extend this analysis to the top 10 U.S. equity ETFs.

Following the discussion in Section 4, I divide the sample into 3 subperiods: the full sample, 2007–2010 and 2009–2010. The regression model is as follows,

$$R_t^{index} = \alpha + \sum_{l=1}^{5} \beta_{R,l} R_{t-l}^{index} + \sum_{l=0}^{5} \beta_{Flow,l} Flow_{t-l} + \varepsilon_t \tag{4}$$

where R_t^{index} is the index return at t, $Flow_{t-1}$ is the flow from equation (1) at $t-l$ with l taking values from zero to five. The inclusion of lags of the dependent and independent variables controls for any autocorrelation within flows and index returns. Moreover, performing regressions on subperiods allows some flexibility in adjusting for time-varying autocorrelation patterns (Lo and Wang 2009). The mutual fund flow literature (Edelen and Warner 2001, Ben-Rephael et al. 2011) also suggests four to five lags for daily regressions.

Table 5 reports individual fund-level regression results for the S&P 500 ETF (SPY) and NASDAQ 100 ETF (QQQ) that track S&P 500 and NASDAQ 100 indices respectively. The estimates on the S&P 500 ETF (SPY) contemporaneous flow coefficient are positive and highly significant (p-value less than 0.01) in all periods. The coefficient estimate on two-day lagged flow for the S&P 500 ETF is negative and statistically significant in the 2007–2010 period. Contemporaneous flow coefficients in the case of the NASDAQ 100 ETF are positive, significant and increasing in magnitude across subperiods.

Table 5. S&P500 and NASDAQ100 ETFs: Fund-Level Regressions

The regression table shows estimation results for flow-return regressions for two ETFs, S&P 500 ETF and NASDAQ-100 ETF which track S&P 500 and NASDAQ 100 indices respectively. The regressions were run in each of the three subperiods: 1993–2010 (full sample), 2007–2010 and 2009–2010. $Flow_t$ and $Flow_{t-1}$ are the contemporaneous and lagged flows from equation (1). Contemporaneous and lagged index returns are denoted as $R_{index,t}$ and $R_{index,t-l}$ with l varying from one to five. Standard errors are corrected for heteroscedasticity and autocorrelation up to five lags with Newey and West (1987) procedure (t-stats are in parenthesis). *, **, *** measure significance at the 10%, 5%, and 1% level, respectively.

	S&P 500 ETF			NASDAQ100 ETF		
Variables	1993–2010	2007–2010	2009–2010	2004–2010	2007–2010	2009–2010
$R_{index,t-1}$	-0.071***	-0.137***	-0.034	-0.111***	-0.138***	-0.006
	(-3.072)	(-2.755)	(-0.512)	(-3.100)	(-3.168)	(-0.108)
$R_{index,t-2}$	-0.067**	-0.123*	-0.046	-0.098**	-0.115*	0.009
	(-2.148)	(-1.882)	(-0.652)	(-1.962)	(-1.867)	(0.140)
$R_{index,t-3}$	0.011	0.053	-0.005	0.049	0.052	-0.042
	(0.462)	(1.086)	(-0.087)	(1.336)	(1.150)	(-0.671)
$R_{index,t-4}$	-0.017	-0.026	0.033	-0.033	-0.036	-0.027
	(-0.620)	(-0.433)	(0.485)	(-0.804)	(-0.680)	(-0.321)
$R_{index,t-5}$	-0.037	-0.020	0.022	0.006	0.014	-0.005
	(-1.192)	(-0.303)	(0.323)	(0.131)	(0.236)	(-0.058)
$flow_t$	0.027***	0.069**	0.218***	0.082***	0.121***	0.172***
	(3.111)	(2.439)	(5.324)	(3.775)	(3.307)	(4.400)
$flow_{t-1}$	0.012	-0.047*	-0.013	0.012	0.035	0.048
	(1.418)	(-1.942)	(-0.322)	(0.330)	(0.546)	(1.090)
$flow_{t-2}$	-0.005	0.023	0.013	-0.008	-0.021	-0.020
	(-0.616)	(0.866)	(0.352)	(-0.299)	(-0.460)	(-0.446)
$flow_{t-3}$	0.005	-0.017	0.040	-0.006	-0.007	-0.021
	(0.640)	(-0.644)	(1.064)	(-0.246)	(-0.196)	(-0.486)
$flow_{t-4}$	-0.007	0.007	-0.012	0.002	-0.005	-0.018
	(-0.956)	(0.251)	(-0.375)	(0.099)	(-0.121)	(-0.376)
$flow_{t-5}$	-0.002	-0.023	-0.049	-0.007	-0.008	0.017
	(-0.299)	(-1.081)	(-1.450)	(-0.278)	(-0.207)	(0.447)
Constant	0.000	-0.000	0.001	0.000	0.000	0.001
	(1.457)	(-0.562)	(0.760)	(0.853)	(0.054)	(0.898)
Observations	4,379	756	252	1,420	756	252
R^2	0.015	0.054	0.175	0.034	0.044	0.070

Numerically, in the case of the S&P 500 ETF, an inflow of one standard deviation ($\sigma_{flow}^{S\&P500} = 2.4\%$) is associated with a return shock of 0.07% during 1993–2010, and shocks of 0.17% and 0.52% for the latter subperiods. Interestingly, given the high kurtosis ($K_{flow,2007-2010} = 18.23$) for the S&P 500 ETF (SPY), the possibility of larger flows is higher than for the normal distribution. For instance, the 90$^{\text{th}}$ percentile of the flow distribution for S&P 500 ETF (SPY) during 2007–2010 corresponds to a 20% change in shares outstanding which implies a return shock of 2%. In monetary terms, a one standard deviation shift in flow is equal to a dollar flow of \$2.11 billion and a 90$^{\text{th}}$ percentile shift in flow is equal to a dollar flow of \$13.2 billion.

To study the flow-return relation further, I run regression (4) on each one the top 10 ETFs from Table 2 and report the coefficients for the first lag for the underlying index return and for the contemporaneous coefficient and first three lags for the flow in Table 6. The data before 2007 is lower quality with long spells of missing observations, especially for the iShares family of funds, hence the starting year varies.

Table 6. Top 10 ETFs: Individual Regressions

The regression table shows fund-level estimation results for top 10 ETFs ranked by total net assets during two subperiods: 2007–2010 and 2009–2010. For iShares Russell 1000 Growth (IWF) and iShares Russell 1000 Growth (IWD) ETFs, shares outstanding data exists starting 2008. $Flow_t$ and $Flow_{t-1}$ are the contemporaneous and lagged flows from equation (1). Contemporaneous and lagged index returns are denoted as $R_{index,t}$ and $R_{index,t-l}$. Style is based on market capitalization of the tracked index components obtained from *Bloomberg Terminal* and can take the following values: LC - Large Cap, MC - Middle Cap, SC - Small Cap. Standard errors are corrected for heteroscedasticity and autocorrelation up to five lags with Newey and West (1987) standard errors (t-stats are in parenthesis). *, **, *** measure significance at the 10%, 5%, and 1% level, respectively.

ETF	Ticker	Style	Period	$R_{index,t-1}$	$flow_t$	$flow_{t-1}$	$flow_{t-2}$	$flow_{t-3}$	R^2
SPDR S&P 500	SPY	LC	2007–2010	−0.137***	0.069**	−0.047*	0.023	−0.017	0.054
				(−2.755)	(2.439)	(−1.942)	(0.866)	(−0.644)	
			2009–2010	−0.034	0.218***	−0.013	0.013	0.040	0.175
				(−0.512)	(5.324)	(−0.322)	(0.352)	(1.064)	
iShares S&P 500	IVV	LC	2007–2010	−0.146***	−0.048	0.015	0.012	0.081***	0.045
				(−2.87)	(−1.38)	(0.26)	(0.33)	(2.67)	
			2009–2010	−0.022	0.443***	−0.045	0.168	−0.048	0.050
				(−0.38)	(3.80)	(−0.37)	(1.56)	(−0.38)	
PowerShares NASDAQ 100	QQQQ	LC	2007–2010	−0.138***	0.121***	0.035	−0.021	−0.007	0.044
				(−3.168)	(3.307)	(0.546)	(−0.460)	(−0.196)	
			2009–2010	−0.006	0.172***	0.048	−0.020	−0.021	0.070
				(−0.108)	(4.400)	(1.090)	(−0.446)	(−0.486)	
Vanguard Total Stock Mkt	VTI	LC	2007–2010	−0.128**	0.070	0.167	0.002	0.034	0.038
				(−2.54)	(0.32)	(1.11)	(0.01)	(0.24)	
			2009–2010	−0.059	0.931**	0.522	−0.766	−0.320	0.059
				(−0.92)	(2.24)	(1.54)	(−1.56)	(−0.95)	
iShares Russell 2000	IWM	SC	2007–2010	−0.107**	0.230***	0.034	−0.077**	−0.020	0.104
				(−1.976)	(6.164)	(0.668)	(−2.399)	(−0.539)	
			2009–2010	0.045	0.219***	−0.023	−0.072**	0.021	0.184
				(0.759)	(6.59)	(−0.75)	(−2.40)	(0.64)	
iShares Russell 1000 Growth	IWF	LC	2008–2010	−0.104**	0.402***	0.273*	−0.654***	0.213**	0.058
				(−2.221)	(3.749)	(1.796)	(−3.779)	(2.520)	

(Continued)

Table 6. – *Continued*

ETF	Ticker	Style	Period	$R_{index,t-1}$	$flow_t$	$flow_{t-1}$	$flow_{t-2}$	$flow_{t-3}$	R^2
			2009–2010	−0.034	0.351	0.072	−0.395	−0.372	0.024
				(−0.57)	(1.05)	(0.29)	(−1.58)	(−1.55)	
iShares Russell 1000 Value	IWD	LC	2008–2010	−0.121**	0.161***	0.096**	−0.306***	0.144***	0.066
				(−2.355)	(4.019)	(2.228)	(−8.353)	(3.873)	
			2009–2010	−0.042	−0.048	−0.617*	−0.100	−0.190	0.025
				(−0.760)	(−0.178)	(−1.752)	(−0.293)	(−0.667)	
SPDR S&P MidCap 400	MDY	MC	2007–2010	−0.084*	0.188***	0.026	−0.048	0.023	0.052
				(−1.65)	(3.71)	(0.63)	(−1.03)	(0.54)	
			2009–2010	0.008	0.254***	0.033	−0.146**	−0.006	0.093
				(0.13)	(4.04)	(0.66)	(−2.21)	(−0.09)	
SPDR DJIA	DIA	LC	2007–2010	−0.146***	−0.016	−0.035	−0.001	0.017	0.047
				(−2.71)	(−0.72)	(−1.59)	(−0.04)	(0.76)	
			2009–2010	−0.052	−0.006	−0.039	−0.058*	−0.031	0.013
				(−0.84)	(−0.19)	(−1.24)	(−1.91)	(−0.91)	
iShares S&P MidCap 400	IJH	MC	2007–2010	−0.082*	0.340***	0.035	0.105	−0.086	0.026
				(−1.664)	(2.622)	(0.274)	(0.823)	(−0.695)	
			2009–2010	0.008	0.556***	0.226	−0.029	−0.038	0.034
				(0.166)	(3.054)	(0.907)	(−0.136)	(−0.178)	

The contemporaneous flow coefficient estimates in Table 6 are significant and positive in any of the subperiods for nine out of ten funds. All of the significant coefficient estimates for the two-day lagged flow display a negative sign, which lends support to the price-reversal hypothesis.

4.3 Panel Analysis

Given that equity index returns are usually correlated, running pooled regressions without accounting for the cross-panel residual correlation is not recommended (Wooldridge 2001). Usually, a regression methodology adjustment is two-way clustering (Petersen 2009) or heteroscedasticity and cross-correlation robust standard errors (Haas and Lelyveld 2006, Kacperczyk and Seru 2007). Therefore, I adopt panel corrected standard errors with adjustments for cross-panel correlation and a panel-specific AR(1) process. Since lagged index return coefficients in Tables 5 and 6 are significant for some periods, estimation of the within-fund autocorrelation should increase the robustness of my inferences. I also reestimate results in this section using Petersen (2009)'s two-way clustering, which leads to no major changes in my findings results. I also apply random and fixed effect panel estimators and obtain quantitatively similar results[28]

In this section, I extend the model used in the individual regression analysis in Section 4.2 to a panel specification with two subsamples, the top 10 ETFs and the entire sample of 286 U.S. equity ETFs, for three subperiods: full sample, 2007–2009, 2009–2010. Again, the regression specification is similar to the one used in Section 4.2 and is as follows,

[28]Beck and Katz (1995) state that feasible generalized least squares (FGLS) used to estimated random effect model with unknown variance provide overly optimistic estimates. However, random effects model is still useful as a comparison. Using random effects model, I obtain larger t-stats for the lagged index returns and slightly lower t-stats for the flow variables which provide same inferences as the panel corrected standard errors.

$$R_{i,j,t}^{index} = \alpha + \sum_{l=1}^{5} \beta_{R,l} R_{j,t-l}^{index} + \sum_{l=0}^{5} \beta_{Flow,l} Flow_{i,j,t-l} + \varepsilon_{i,t} \tag{5}$$

where $R_{j,t}^{index}$ corresponds to the daily return of the j^{th} index tracked by the i^{th} fund at t , $R_{j,t-l}^{index}$ is the l-lagged j^{th} index return, $Flow_{i,t-l}$ is the l-lagged flow of the i^{th} fund which tracks the j^{th} index. The lag indicator l varies between 0 to 5 for the flows and between 1 and 5 for the index returns. In the sample, by definition $i \geq j$ as there are more funds than indices.

The coefficients for the contemporaneous ETF flow are positive and strongly significant for all subsamples and subperiods (with t-stats ranging from 4.313 to 9.249). The coefficients for the lagged flows are significantly negative (t-stats of -1.864) at the second lag for top 10 ETFs for the subperiods of 2007–2010 and 2009–2010. Interestingly, at lags four and five, flow coefficients for the whole sample and across subperiods become negative and significant as well (with absolute t-stats higher than 2). In other words, lagged flows tend to be negatively correlated with returns not only for the top 10 ETFs, but for the whole sample as well, which is consistent with the price-reversal hypothesis.

For the 2007–2010 period, a one standard deviation shift in flows for the top 10 ETFs is about 2%, which is associated with an 18 basis points (bp) increase in the corresponding index return. For the 2009–2010 period, the magnitude of the relation is about two times larger with a corresponding return increase of 36 bp. The full sample analysis panel in Table 7 reports a somewhat smaller magnitude of the contemporaneous flow-return relation. Specifically, a one standard deviation shifts in total flows in 2007–2010 ($\sigma_{flow} = 3.3\%$) and in 2009–2010 ($\sigma_{flow} = 2.5\%$) are related to an increase in returns of five and four basis points respectively.

It is important to note that panel tests based on the full sample have low explanatory power as the vast majority of funds have low total net assets and relatively infrequent

Table 7. Panel Regressions

The regression table reports estimation results for panel regressions of index return on flow for several subsamples and subperiods. First subsample contains top 10 ETFs ranked by net assets and second subsample is the whole U.S. equity ETFs sample. The estimation is conducted for three subperiods: 1993–2010, 2007–2010 and 2009–2010. $Flow_t$ and $Flow_{t-l}$ are the contemporaneous and l-lagged flows from equation (1). Contemporaneous and l-lagged index returns are denoted as $R_{index,t}$ and $R_{index,t-l}$. Standard errors are Panel Corrected Standard Errors (Kacperczyk and Seru 2007) corrected for heteroscedasticity, cross-panel correlation and panel-specific autocorrelation specified as AR(1) process (z-stats are in parenthesis). *, **, *** measure significance at the 10%, 5%, and 1% level, respectively.

	Top 10 ETFs			Full Sample		
Variables	1993–2010	2007–2010	2009–2010	1993–2010	2007–2010	2009–2010
$R_{index,t-1}$	-0.065***	-0.119***	-0.011	-0.075***	-0.105***	-0.013
	(-3.850)	(-3.448)	(-0.188)	(-3.634)	(-3.535)	(-0.233)
$R_{index,t-2}$	-0.070***	-0.079**	-0.019	-0.071***	-0.078***	-0.032
	(-4.138)	(-2.284)	(-0.319)	(-3.462)	(-2.622)	(-0.593)
$R_{index,t-3}$	0.031*	0.051	-0.018	0.027	0.031	-0.004
	(1.835)	(1.467)	(-0.303)	(1.302)	(1.033)	(-0.083)
$R_{index,t-4}$	-0.017	-0.035	0.001	-0.023	-0.031	-0.000
	(-1.020)	(-1.004)	(0.019)	(-1.118)	(-1.044)	(-0.003)
$R_{index,t-5}$	-0.030*	-0.021	-0.003	-0.031	-0.031	-0.007
	(-1.786)	(-0.613)	(-0.052)	(-1.515)	(-1.052)	(-0.139)
$Flow_t$	0.018***	0.093***	0.179***	0.010***	0.014***	0.016***
	(4.427)	(6.774)	(9.249)	(7.432)	(5.202)	(4.313)
$Flow_{t-1}$	-0.001	0.006	0.012	0.001	-0.000	-0.001
	(-0.257)	(0.431)	(0.695)	(0.454)	(-0.172)	(-0.185)
$Flow_{t-2}$	-0.005	-0.025*	-0.040**	-0.000	-0.002	-0.005
	(-1.399)	(-1.864)	(-2.297)	(-0.053)	(-0.957)	(-1.276)
$Flow_{t-3}$	0.001	0.012	0.004	0.000	0.001	0.000
	(0.271)	(0.928)	(0.228)	(0.313)	(0.558)	(0.102)
$Flow_{t-4}$	-0.004	0.006	-0.014	-0.002	-0.004	-0.004
	(-1.188)	(0.489)	(-0.779)	(-1.628)	(-1.490)	(-1.247)
$Flow_{t-5}$	0.005	-0.004	-0.002	-0.002**	-0.006**	0.001
	(1.455)	(-0.330)	(-0.126)	(-2.006)	(-2.530)	(0.198)
Constant	0.000	-0.000	0.001	0.000	-0.000	0.001
	(0.661)	(-0.425)	(0.864)	(0.432)	(-0.107)	(0.923)
Obs	27,668	7,541	2,520	387,879	194,809	70,335
R^2	0.013	0.033	0.050	0.013	0.020	0.002
Panels	10	10	10	286	286	285

creation-redemption activity. Hence, there is a high ratio of observations with zero net flows. For instance, during 2009, the top 10 ETFs had zero net flow activity during 20% of the total trading days, while the remaining 276 ETFs had no net flow activity during 77% of the days. In these conditions, tests based on a larger sample will necessarily have a downward bias. I also test other specifications for the regression equation (5), including varying index return and flow lags and obtain quantitatively similar results.

4.4 Aggregate Flow

4.4.1 Multiple Funds per Index

Given that an index can be tracked by several ETFs with similar features (e.g., expenses), relatively high correlations in flows between these funds could imply a common factor driving those flows. Both the information and return chasing hypothesis imply that flows across funds that share the same features and track the same index should be correlated, as investors react either to information shocks or to past returns. Low correlations, on the other hand, suggest that there is less probability of a common factor driving flows.

The most tracked index is S&P 500 (SPX) used by nine ETFs (equity and leveraged). The 2$^{\text{nd}}$ most tracked index is NASDAQ 100 (NDX), which is used as reference by six ETFs, five of which are leveraged and only one unleveraged. In the sample, there are seven indices tracked by two ETFs, while the remaining 264 indices are tracked by one fund each.

If several funds track the same underlying asset basket, then flows to these funds entail APs buying or selling the same asset basket. If the flows are driven by the factors correlated with the underlying stocks, then flows to different funds tracking the same stocks should be highly and positively correlated on average. Return chasing (Goetzmann and Massa 2002) or information arrival (Edelen and Warner 2001) could play the role of a common factor. In this case, a positive association between flows and returns is more likely to be due to a common factor affecting flows and returns simultaneously instead of the price pressure exerted by the fund flows. Contrarily, if the flows are less subject to the common factor and are instead generated exogenously, then correlations should be low on average.

Table 8. Correlations of Flows of Multiple Funds Tracking the Same Index

The table shows auto and cross-correlations of fund flows tracking the same index. $FFlow$, $Flow$ and $LFlow$ are lead, contemporaneous and lagged flows respectively. Subindex denotes the fund ticker. Panel A presents auto and cross-correlations for flows for SPDR S&P 500 ETF (SPY) and iShares S&P 500 (IVV) which track the same index S&P500. Panel B presents auto and cross-correlations for flows for iShares S&P 400 ETF (IJH) and SPDR S&P MidCap 400 (MDY) which track the same index S&P MidCap 400. The cross-correlation period is the shortest common period to the pair which for Panel A is from 2000 to 2010 and for Panel B from 2000 to 2010. The auto correlation uses full available data period for each fund. The asterisks *, **, *** represent statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A

	$FFlow_{SPY}$	$FFlow_{IVV}$	$Flow_{SPY}$	$Flow_{IVV}$	$LFlow_{SPY}$
$FFlow_{SPY}$	1				
$FFlow_{IVV}$	-0.012	1			
$Flow_{SPY}$	-0.073***	0.099***	1		
$Flow_{IVV}$	-0.004	0.007	-0.012	1	
$LFlow_{SPY}$	0.028	-0.041*	-0.073***	0.099***	1
$LFlow_{IVV}$	0.028	-0.006	-0.004	0.007	-0.012

Panel B

	$FFlow_{IJH}$	$FFlow_{MDY}$	$Flow_{IJH}$	$Flow_{MDY}$	$LFlow_{IJH}$
$FFlow_{IJH}$	1				
$FFlow_{MDY}$	-0.015	1			
$Flow_{IJH}$	-0.029	0.035	1		
$Flow_{MDY}$	0.026	-0.015	-0.015	1	
$LFlow_{IJH}$	0.053**	0.035	-0.029	0.035	1
$LFlow_{MDY}$	-0.01	-0.002	0.026	-0.015	-0.015

Table 8 presents auto and cross-correlations of fund flows for four ETFs tracking two indices, the S&P 500 and the S&P MidCap 400[29]. If flows are generated by a common factor, the contemporaneous correlation between flows should be relatively high, however the evidence in Table 8 seems to indicate otherwise. Contemporaneous correlation for the SPY-IVV and IJH-MDY fund pairs in Panels A and B is -0.012 and -0.015 respectively, both statistically indifferent from zero. Restricting the sample period to 2007–2010, slightly reduces correlations with $\rho^t_{SPY,IVV}$ falling to -0.04 and $\rho^t_{SPY,IVV}$ decreasing to -0.077 and becoming significant with p-value of 0.035. There are no substantial differences in expense fees or other characteristics between the funds that would imply such a

[29]Extending the same analysis to the other four indices tracked by two ETFs each provides similar conclusions. Average contemporaneous correlation between the same-index flows is 0.04 and insignificantly different from zero.

differential response to a common factor if such factor existed. Moreover, both panels feature funds in the top 10 ETFs ranked by net assets with relatively high creation-redemption activity; hence, non-zero flow observations are not an issue.

Autocorrelations in Table 8 exhibit varied patterns. In the case of the S&P 500 ETF, autocorrelation is slightly negative ($\rho_{SPY}^{t-1,t}$=-0.073) and highly significant, but the same pattern fails to appear for the other ETF tracking the same index or for the other pair of funds tracking S&P MidCap 400.

4.4.2 Aggregate Equity ETF Flow

Mutual fund researchers usually test flow-return hypotheses using aggregate flows given that high-frequency fund-level flows data are quite rare[30]. Moreover, high-frequency aggregate data, even if available, is also often constrained to a certain period. For instance, length of the sample period is two years in Ben-Rephael et al. (2011), four years in Goetzmann and Massa (2003) and one year in Edelen and Warner (2001). In the case of ETFs, the available cross-section of funds is usually close to the total funds in the economy. Consequently, testing Hypothesis 1 on the daily aggregate level allows me to conduct a direct comparison with the mutual fund flow literature findings.

In this section, I use two measures for aggregate ETF flow based on the mutual fund literature. The first measure, used by Edelen and Warner (2001), is essentially a percentage growth in assets under management, less the percentage change in NAV. I employ a similar formulation and use the previous day NAVs when calculating both previous and contemporaneous net assets with the following specification,

$$Flow_t^{aggr} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n}(Shrout_{i,j,t}NAV_{i,j,t})}{\sum_{j=1}^{k}\sum_{i=1}^{n}(Shrout_{i,j,t-1}NAV_{i,j,t})} - 1 \tag{6}$$

[30]A notable exception is the paper by Rakowski and Wang (2009) where the authors have access to the daily mutual fund data from Lipper covering 6,772 funds

where $Shrout_{i,j,t}$ and $NAV_{i,j,t}$ are the shares outstanding and net asset value respectively for fund i tracking index j at time t, n is the number of funds tracking index j at time t, and k is the total number of indices. Using contemporaneous NAVs to determine assets under management for both days helps to avoid implicitly including fund returns in aggregate flow.

The second measure is currency flow, which is also used in the mutual fund literature (e.g., Ben-Rephael et al. (2011)). This measure represents an increase in assets under management in corresponding currency units. Aggregate dollar ETF flows are economically significant with a standard deviation of \$2.47 billion during 2007–2010, which is comparable to mutual fund flows[31]. I use the dollar flow definition from equation (3) and sum dollar flows across ETFs at time t,

$$DFlow_t^{aggr} = \sum_{j=1}^{k}\sum_{i=1}^{n}(Shrout_{i,j,t} - Shrout_{i,j.t-1})NAV_{i,j,t} \tag{7}$$

where the notation is the same as in equation (6).

The regression of index returns on the aggregate ETF flows uses a similar approach to the Section 4.3 with the specification as follows,

$$R_t^{index} = \alpha + \sum_{l=1}^{5}\beta_{R,l}R_{t-l}^{index} + \sum_{l=0}^{5}\beta_{Flow^{aggr},l}Flow_{t-l}^{aggr} + \varepsilon_t \tag{8}$$

where R_t^{index} and R_{t-l}^{index} is the contemporaneous and l-lagged index return at time t, $Flow_{t-l}^{aggr}$ is the aggregate flow for U.S. equity ETFs at time t with l-varying lag using either flow from equation (6) or dollar flow from equation (7). The lag indicator, l, varies from one to five for the index return and from zero to five for the aggregate dollar flow. Both flow and dollar flow regressions are performed on all subperiods, but only the 2007–

[31]See, for example, *ICI Weekly Estimated Long-Term Mutual Fund Flows*, `http://www.ici.org/info/flows_data_2013.xls`

2010 subperiod dollar flow is reported as the results across subperiods are quantitatively similar.

The aggregate flow-return studies usually have some freedom in choosing the underlying return indices that are most appropriate for the tested hypothesis. For instance, Goetzmann and Massa (2003) use the S&P 500 Index, Edelen and Warner (2001) use the NYSE Index and Ben-Rephael et al. (2011) use the TASE-25 Index. I report regression results using two indices as the dependent variables in regressions in this section: Value-Weighted Index Return from CRSP and the S&P 500 Index.

Table 9 describes the regression results of returns on aggregate flow. Regressions using scaled flows (column $Flow$) from Edelen and Warner (2001) exhibit positive and significant contemporaneous coefficients for $flow_t$ with t-stats from 3.305 to 6.102, which lends support to Hypothesis 1. The first lag exhibits an approximately four times smaller but still positive coefficient with t-stats of 1.75 and 1.82 for the VWRETD and S&P 500 index returns respectively. Lags two and four, on the other hand, are negative and statistically significant in the 1993–2010 subperiod with t-stats ranging from 1.88 to 2.05. This implies price reversal from the contemporaneous flow-return shock. In terms of economic impact, I find that a one standard deviation flow shock ($\sigma^{aggr}_{flow} = 0.011$ for 2007–2010 and 0.006 for 2009–2010) is associated on average with a 52 bp return shock in the market return for the 2007–2010 period and with a 54 bp return shock for the 2009–2010 period.

For the aggregate dollar flow regression (column $DFlow$), coefficient estimates for contemporaneous flows, $DFlow_t$, for both indices are positive and significant. A one standard deviation shock of \$2.47 billion is associated with 50 basis points shock in returns. Moreover, aggregate dollar flows maintain significant, albeit slightly reduced, kurtosis (K=11.26). Hence, larger shocks than warranted by the normal distribution have a higher possibility of occurring. For instance, the 90$^{\text{th}}$ percentile of the flow distribution corresponds to a positive inflow of \$14.42 billion which implies an associated return shock

Table 9. Aggregate Sample Flow Regressions

This table present aggregate ETF return on flow regressions for three subperiods and two different choices of underlying index: Value-Weighted Index Return (CRSP) and S&P 500 Index. Regressions were estimated with equation (8). In the *Flow* columns, $Flow_t^{aggr}$ and $Flow_{t-l}^{aggr}$ are contemporaneous and lagged scaled dollar ETF flows described by equation (6) and similar to Edelen and Warner (2001) flow definition. In the *DFlow* columns, $Flow_t^{aggr}$ and $Flow_{t-l}^{aggr}$ are the contemporaneous and lagged aggregated dollar ETF flows expressed in $ billions from equation (7). The lag operator l varies from zero to five. Contemporaneous and lagged index returns are denoted as $R_{index,t}$ and $R_{index,t-l}$ with l varying from one to five. For the VWRETD columns, $R_{index,t}$ is Value-Weighted Index Return (CRSP) and for S&P 500 columns, $R_{index,t}$ is the S&P 500 index return. Standard errors are corrected for heteroscedasticity and autocorrelation up to five lags with Newey-West (t-stats are in parenthesis). The asterisks *, **, *** represent statistical significance at the 10%, 5%, and 1% level, respectively.

	VWRETD				S&P 500			
	Flow	Flow	Flow	DFlow	Flow	Flow	Flow	DFlow
Variables	1993–2010	2007–2010	2009–2010	2007–2010	1993–2010	2007–2010	2009–2010	2007–2010
$R_{index,t-1}$	−0.035	-0.101**	-0.041	-0.105**	-0.073***	-0.141***	-0.072	-0.144***
	(−1.483)	(-2.056)	(-0.631)	(-2.113)	(-3.076)	(-2.811)	(-1.077)	(-2.815)
$R_{index,t-2}$	−0.056*	-0.041	-0.032	-0.090	-0.066**	-0.065	-0.033	-0.114*
	(−1.877)	(-0.761)	(-0.424)	(-1.346)	(-2.229)	(-1.173)	(-0.448)	(-1.708)
$R_{index,t-3}$	0.026	0.053	-0.011	0.047	0.016	0.057	-0.007	0.047
	(1.036)	(0.969)	(-0.200)	(0.919)	(0.648)	(1.055)	(-0.123)	(0.930)
$R_{index,t-4}$	−0.012	0.001	0.050	-0.013	-0.016	0.003	0.045	-0.014
	(−0.411)	(0.014)	(0.710)	(-0.223)	(-0.551)	(0.045)	(0.652)	(-0.228)
$R_{index,t-5}$	−0.040	-0.009	0.004	-0.009	-0.039	-0.004	0.022	-0.004
	(−1.307)	(-0.143)	(0.055)	(-0.141)	(-1.275)	(-0.055)	(0.317)	(-0.057)
$flow_t$	0.106***	0.464***	0.963***	0.002***	0.111***	0.457***	0.921***	0.002***
	(3.305)	(5.922)	(6.102)	(5.414)	(3.435)	(5.758)	(5.945)	(5.303)
$flow_{t-1}$	0.025	-0.023	0.250*	-0.000	0.029	-0.002	0.249*	-0.000
	(1.312)	(-0.449)	(1.745)	(-1.133)	(1.603)	(-0.038)	(1.820)	(-1.048)
$flow_{t-2}$	−0.039*	-0.179	-0.091	-0.000	-0.040**	-0.171	-0.092	-0.000
	(−1.881)	(-1.524)	(-0.769)	(-0.267)	(-2.032)	(-1.468)	(-0.770)	(-0.168)
$flow_{t-3}$	0.006	0.030	-0.001	-0.000	0.009	0.027	-0.010	-0.000
	(0.473)	(0.443)	(-0.008)	(-0.025)	(0.671)	(0.396)	(-0.073)	(-0.095)
$flow_{t-4}$	−0.024**	-0.025	-0.149	-0.000	-0.021*	-0.025	-0.137	-0.000
	(−2.045)	(-0.407)	(-1.271)	(-0.455)	(-1.916)	(-0.387)	(-1.217)	(-0.283)
$flow_{t-5}$	0.028**	0.035	-0.192	-0.000	0.029**	0.047	-0.196	-0.000
	(2.150)	(0.358)	(-1.383)	(-1.594)	(2.150)	(0.487)	(-1.436)	(-1.451)
Constant	0.000	-0.000	0.000	-0.000	0.000	-0.001	0.000	-0.000
	(1.331)	(-0.553)	(0.687)	(-0.361)	(0.888)	(-0.849)	(0.551)	(-0.658)
Obs	3.887	751	248	751	3,887	751	248	751
R^2	0.025	0.107	0.217	0.073	0.030	0.116	0.274	0.083

of 4%. Furthermore, dollar flow coefficients at lags four and five exhibit negative signs and are weakly significant with t-stats from -1.594 to -0.455.

Overall, the time series regression evidence seems to strongly support the contemporaneous price pressure hypothesis while providing considerable support for the price reversal hypothesis.

4.4.3 VAR Analysis of Aggregate Flows and Returns

While the panel and aggregate flow regressions provide some evidence for the transitory price pressure hypothesis, we can learn more about the flow-return structure over time by studying it in a vector autoregressive (VAR) framework. VAR models are generally used to measure dynamic relation between several variables (Hamilton 1994) and, in my case, to capture interdependencies between flows and returns at various lags and to isolate the price reversal effect if present. Furthermore, VAR models are well known in the institutional flows and returns literature. For instance, Froot et al. (2001) have applied VAR to portfolio flows in international context, while Watson and Wickramanayake (2012), Oh and Parwada (2007), Ben-David et al. (2011) study the mutual fund flow-return relation using VAR in Korea, Australia and Israel respectively. Rakowski and Wang (2009) employ VAR to study short-term U.S. mutual fund flows. Interestingly, out of the four studies mentioned, only Ben-David et al. (2011) find significant evidence of price pressure.

To maintain comparability across approaches and previous sections, the period from 2007 to 2010 comprising 756 trading days is selected. This sample size is similar to other short-term mutual fund flow studies. The VAR model is specified as follows,

$$\begin{bmatrix} R_t^{index} \\ Flow_t^{aggr} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \beta_{1R}(L) & \beta_{1Flow^{aggr}}(L) \\ \beta_{2R}(L) & \beta_{2Flow^{aggr}}(L) \end{bmatrix} \begin{bmatrix} R_{t-l}^{index} \\ Flow_{t-l}^{aggr} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{9}$$

with L determined by the Akaike criterion ($AIC_{min} = -11.3$) and equal to four[32]. The contemporaneous shock and four lags represent the entire trading week and are comparable to the panel and aggregate regressions in Section 4.3. Furthermore, lag order selection test results are similar to Ben-Rephael et al. (2011), Edwards and Zhang (1998) who also use four lags in their VAR models. In the estimation results, for convenience, I denote the regression with R_t^{index} as the dependent variable as Equation 1 and the regression with $Flow_t^{aggr}$ as the dependent variable as Equation 2 .

The results of the VAR model estimation are presented in Table 10. The lagged aggregate flow coefficient estimates for Equation 1, $\beta_{1Flow^{aggr}}$, exhibit generally negative and significant or insignificantly positive values with the two-day lagged coefficient estimate's magnitude of -0.160 with a *t*-stat of -2.486 (*p*-value=0.013). The return coefficient estimates for the same equation, β_{1R}, are significantly negative for lags one and two, which indicates a negative serial autocorrelation in returns. In Equation 2, lagged flow coefficient estimates do not exhibit statistical significance, while lagged returns are negatively related to the contemporaneous flows.

Cumulative impulse response (CIRF) functions for the VAR model in equation (9) are illustrated in Figure 4. Graph 4a describes the response of a one unit flow shock on index returns. The contemporaneous flow-return shock with magnitude of 0.462 (*t*-stat of 7.65) is placed at lag one in the graphs. During day one (lag two in the graphs), this contemporaneous shock decreases slightly to 0.409. This decrease in CIRF during the day one after the shock is statistically insignificant (*t*-stat of -0.65) and in other VAR specifications changes sign to positive. The insignificance of the lag one impulse response function (IRF) most likely signals a combination of two countervailing forces: price pressure and consequent price reversal. On day two, the accumulated price impact falls to 0.219 with a highly significant difference of 0.243 (*t*-stat of 2.52, at $p\text{-value} < 0.01$) between the contemporaneous shock and the day two cumulative impact. In other

[32]For more information on lag order selection, please see Tsay (2010)

Table 10. Aggregate Flow VAR Regressions

This table present vector autoregressive (VAR) regressions of aggregate ETF flow ($Flow^{aggr}$) and index return (R_{index}) using specification (9). The estimation period is from 2007 to 2010. Aggregate ETF flow is specified as percentage change in aggregate assets under management using contemporaneous NAV as referenced in equation (6). Index return is Value-Weighted Index Return (CRSP). Lagged flows are denoted as $Flow^{aggr}_{t-l}$ with l varying from one to four. Lagged index returns are denoted as $R_{index,t-l}$ with l varying from one to four. The t-stats are in parenthesis. The asterisks *, **, *** represent statistical significance at the 10%, 5%, and 1% level, respectively.

Variables	Equation 1: Dep Var = $R_{VWRETD,t}$	Equation 2: Dep Var = $Flow_t$
$Flow_{t-1}$	0.001	0.041
	(0.015)	(1.109)
$Flow_{t-2}$	−0.160**	0.033
	(−2.486)	(0.884)
$Flow_{t-3}$	0.044	0.023
	(0.686)	(0.603)
$Flow_{t-4}$	−0.013	0.013
	(−0.199)	(0.343)
$R_{index,t-1}$	−0.104***	−0.007
	(−2.755)	(−0.301)
$R_{index,t-2}$	−0.076**	−0.081***
	(−2.016)	(−3.710)
$R_{index,t-3}$	0.053	0.002
	(1.409)	(0.070)
$R_{index,t-4}$	−0.040	−0.089***
	(−1.051)	(−4.042)
Constant	−0.000	0.001*
	(−0.103)	(1.894)
Obs	756	756

words, approximately 53% of the contemporaneous price impact is reversed by day two. On day three, the cumulative impact increases to 0.302 with the difference between contemporaneous shock and day three cumulative impact being 0.160 (t-stat of -1.46). In the following days, cumulative price impact stabilizes at 0.28 with the differential between contemporaneous shock and long-run equilibrium at 0.178 (t-stat of 1.71) which is equivalent to 38% of the day zero impact. Given partial reversion, in the long run, 62% and 38% of the initial shock are permanent and transitory respectively lending support to Hypothesis 1.

Graph 4b presents the response of aggregate flow to one unit shock in aggregate flow, or the persistence in flows. The long-run cumulative response is 1.030; however, it is insignificantly different from one with the a t-stat of 0.38, thereby indicating very weak shock persistence. The response of aggregate flows to one unit shock in index returns is depicted in Graph 4c. Flows respond negatively and significantly to the contemporaneous shock in index returns with the long-run equilibrium achieved at -0.213 with a t-stat of -4.38. In Graph 4d, the response of the index return to one unit shock in index returns is plotted demonstrating a negative serial correlation between contemporaneous and lagged returns. Over a period of 20 days, the response to the shock reaches a maximum of roughly 0.835 at days two and five and then stabilizes at a cumulative response of 0.860 with a t-stat of 2.15.

As a robustness check, I estimate orthogonalized impulse response functions (Hamilton 1994) using the same VAR specification as in equation (9) and obtain results closely matching those in Figure 4 and Table 10. Furthermore, I also test a ranges of four to ten lags in VAR and four to twenty periods in the impulse response functions with qualitatively similar results[33]. Additionally, employing aggregate dollar flow from equation (8) instead of aggregate flow lead to the same inferences.

5 Conclusions

Using a unique database of fund-level flows and daily index returns for the sample of 286 U.S. equity ETFs, I find that exchange-traded fund flows exhibit a statistically significant and cross-sectionally consistent positive association with contemporaneous underlying index returns across several specifications, subperiods and aggregation levels. The magnitude of the relation as a response to one standard deviation flow shock varies in different specification and subsamples from 7 to 52 basis points, with coefficient estimates increasing in the latter part of the sample. I find substantial evidence of price reversal in the panel regressions for top 10 ETFs and in vector autoregressive (VAR) model estimation

[33]With a high number of lags in VAR and a low number of periods, the CIRF graphs take longer to converge to a steady state, but the shape and confidence interval bounds of the impulse function in the first four periods are similar to the ones presented in Figure 4

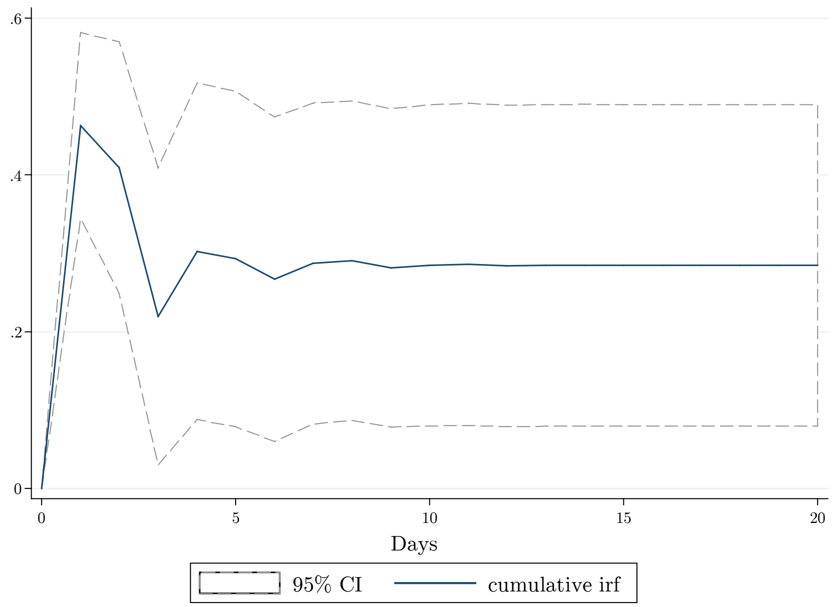
with price reversal accounting for 38% of the contemporaneous return shock due to flows. Given that transitory price pressure is identified by the subsequent price reversal (Ben-Rephael et al. 2011, Goetzmann and Massa 2003), this significant negative relation between lagged fund flow and returns is consistent with the price pressure hypothesis. These findings provide a novel contribution to the price pressure (Harris and Gurel 1986) and institutional flow literature (Edelen and Warner 2001, Goetzmann and Massa 2003) by exploring relation between the fund flows and the underlying returns in the ETF context.

These results are in sharp contrast to the earlier findings by Kalaycıoğlu (2004) who did not find any consistent significant relation between daily flows and returns in a smaller ETF sample. However, their results might be due to the timing of the flow reporting issue (Quinn et al. 2006) prevalent in mutual fund industry (Rakowski and Wang 2009). This bias is difficult to adjust for as flow reporting is voluntary and the data are not verified before publication.

Furthermore, few studies have explored the cross-section of institutional flows as mostly aggregate flows are available. I study flow relation between different funds tracking the same index and find that flows do not seem to be generated by a common factor, such as sentiment (Barberis et al. 1998) or style (Barberis and Shleifer 2003) and are instead more likely to be generated by exogenous factors. Additionally, the common argument that flows are driven by arbitrage activity, which is determined by the fund spread (Agapova 2010, Petajisto 2011), does not seem to hold empirically, as flows are not correlated with the daily spread. However, daily spread is an imperfect estimate of the intraday spread which ultimately defines arbitrage opportunities; hence, examination of the relation between intraday returns and flows could provide additional tests for the price pressure hypothesis.

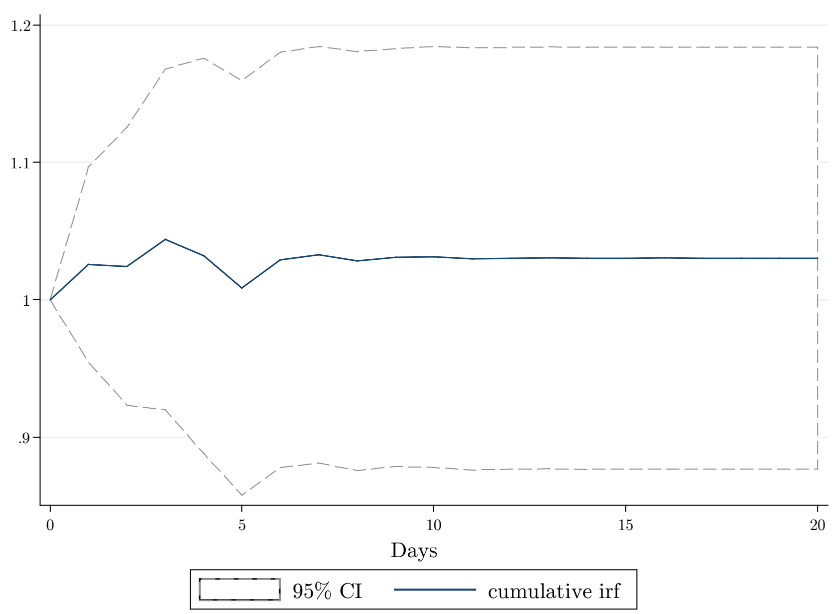
Interestingly, Subrahmanyam (2008), Kalaycıoğlu (2004) and Warther (1995) suggest that the flow-return relation may persist over longer time horizons, from weekly to quarterly. To test if the price pressure effect is not confined to the daily but instead has long-term repercussions, the methodology presented in this paper can be easily extended to longer time horizons using weekly and monthly cross sectional regressions (Subrahmanyam 2009).

Figure 4. Cumulative Impulse Response Functions



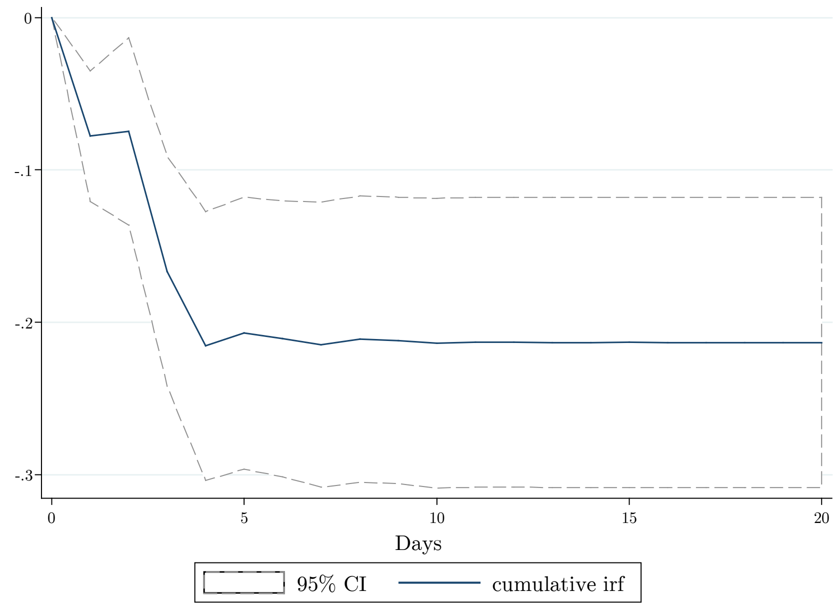
(a) Response of Index Return to One Unit Innovation in Aggregate Flow

Graph 4a outlines the cumulative impulse response of index return (Value-Weighted Index Return) to one unit shock in aggregate flow as measured in equation (6). The contemporaneous relation between index return and aggregate flow is calculated with index return depending on flows specification and is located at lag one. There are 20 lags (days). Confidence intervals of 95% are shown in dashed grey lines.



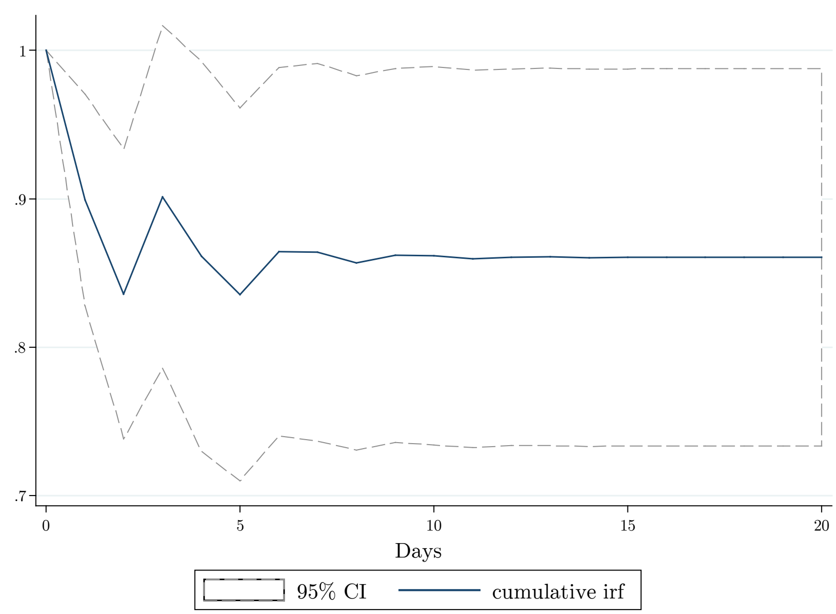
(b) Response of Aggregate Flow to One Unit Innovation in Aggregate Flow

Graph 4b outlines the cumulative impulse response of aggregate flow to one unit shock in aggregate flow as measured in equation (6). The contemporaneous relation between index return and aggregate flow is calculated with index return depending on flows specification and is located at lag one. There are 20 lags. Confidence intervals of 95% are shown in dashed grey lines.



(c) Response of Aggregate Flow to One Unit Innovation in Index Return

Graph 4c outlines the cumulative impulse response of aggregate flow to one unit shock in index return where aggregate flow is described in equation (6). The contemporaneous relation between index return and aggregate flow is calculated with index return depending on flows specification and is located at lag one. There are 20 lags. Confidence intervals of 95% are shown in dashed grey lines.



(d) Response of Index Return to One Unit Innovation in Index Return

Graph 4d outlines the cumulative impulse response of index return to one unit shock in index return. The contemporaneous relation between index return and aggregate flow is calculated with index return depending on flows specification and is located at lag one. There are 20 lags. Confidence intervals of 95% are shown in dashed grey lines.

References

Abner, David J., 2011, *The ETF Handbook: How to Value and Trade Exchange-Traded Funds.*

Ackert, Lucy F. and Yisong S. Tian, 2008, "Arbitrage, Liquidity, and the Valuation of Exchange Traded Funds", *Financial Markets, Institutions & Instruments* **17**.5, pp. 331–362.

Agapova, Anna, 2010, "Conventional mutual index funds versus exchange-traded funds", *Journal of Financial Markets* **In Press, Corrected Proof**.

Barberis, Nicholas and Andrei Shleifer, 2003, "Style investing", *Journal of Financial Economics* **68**.2, pp. 161–199.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, "A model of investor sentiment", *Journal of Financial Economics* **49**.3, pp. 307–343.

Beck, Nathaniel and Jonathan N. Katz, 1995, "What To Do (and Not to Do) with Time-Series Cross-Section Data", *American Political Science Review* **89**.03, pp. 634–647.

Ben-David, Itzhak, Francesco A. Franzoni, and Rabih Moussawi, 2011, "ETFs, Arbitrage, and Contagion", *SSRN eLibrary.*

Ben-Rephael, Azi, S. Kandel, and A. Wohl, 2011, "The Price Pressure of Aggregate Mutual Fund Flows", *Journal of Financial and Quantitative Analysis* **46**.2, p. 585.

Biais, Bruno, Larry Glosten, and Chester Spatt, 2005, "Market microstructure: A survey of microfoundations, empirical results, and policy implications", *Journal of Financial Markets* **8**.2, pp. 217–264.

Boehmer, Ekkehart and J. (Julie) Wu, 2008, "Order Flow and Prices", *SSRN eLibrary.*

Boyer, Brian and Lu Zheng, 2009, "Investor flows and stock market returns", *Journal of Empirical Finance* **16**.1, pp. 87–100.

Cai, Fang and Lu Zheng, 2004, "Institutional trading and stock returns", *Finance Research Letters* **1**.3, pp. 178–189.

Cao, Charles, Eric C. Chang, and Ying Wang, 2008, "An empirical analysis of the dynamic relationship between mutual fund flow and market return volatility", *Journal of Banking & Finance* **32**.10, pp. 2111–2123.

Chan, Louis K. C. and Josef Lakonishok, 1993, "Institutional trades and intraday stock price behavior", *Journal of Financial Economics* **33**.2, pp. 173–199.

Charupat, Narat and Peter Miu, 2011, "The pricing and performance of leveraged exchange-traded funds", *Journal of Banking & Finance* **35**.4, pp. 966–977.

Chen, Guang and Timothy S. Strother, 2008, "On the Contribution of Index Exchange Traded Funds to Price Discovery in the Presence of Price Limits Without Short Selling", *SSRN eLibrary*.

Chen, Constantine Koutsantony, Cameron Truong, and Madhu Veeraraghavan, 2013, "Stock price response to S&P 500 index inclusions: Do options listings and options trading volume matter?", *Journal of International Financial Markets, Institutions and Money* **23**, pp. 379–401.

Chen, Gregory Noronha, and Vijay Singal, 2004, "The Price Response to S&P 500 Index Additions and Deletions: Evidence of Asymmetry and a New Explanation", *The Journal of Finance* **59**.4, pp. 1901–1930.

Delcoure, Natalya and Maosen Zhong, 2007, "On the premiums of iShares", *Journal of Empirical Finance* **14**.2, pp. 168–195.

Denis, Diane K., John J. McConnell, Alexei V. Ovtchinnikov, and Yun Yu, 2003, "S&P 500 Index Additions and Earnings Expectations", *The Journal of Finance* **58**.5, pp. 1821–1840.

Doran, James S., Vaneesha Boney, and David R. Peterson, 2006, "The Effect of the Spider Exchange Traded Fund on the Cash Flow of Funds of S&P Index Mutual Funds", *SSRN eLibrary*.

Edelen, Roger M. and Jerold B. Warner, 2001, "Aggregate price effects of institutional trading: a study of mutual fund flow and market returns", *Journal of Financial Economics* **59**.2, pp. 195–220.

Edwards, Franklin R. and Xin Zhang, 1998, "Mutual Funds and Stock and Bond Market Stability", *Journal of Financial Services Research* **13**.3, p. 257.

Froot, Kenneth A., Paul G.J. O'Connell, and Mark S. Seasholes, 2001, "The portfolio flows of international investors", *Journal of Financial Economics* **59**.2, pp. 151–193.

Goetzmann, William N. and Massimo Massa, 2002, "Daily Momentum and Contrarian Behavior of Index Fund Investors", *Journal of Financial and Quantitative Analysis* **37**.03, pp. 375–389.

Goetzmann, William N. and Massimo Massa, 2003, "Index Funds and Stock Market Growth", *The Journal of Business* **76**.1, pp. 1–28.

Gorton, Gary B. and George G. Pennacchi, 1993, "Security Baskets and Index-Linked Securities", *The Journal of Business* **66**.1, pp. 1–27.

Haas, Ralph de and Iman van Lelyveld, 2006, "Foreign banks and credit stability in Central and Eastern Europe. A panel data analysis", *Journal of Banking & Finance* **30**.7, pp. 1927–1952.

Hamilton, James D., 1994, *The Time Series Analysis*, Princeton University Press.

Hamm, S., 2010, "The Effect of ETFs on Stock Liquidity", *SSRN eLibrary*.

Harris, Lawrence and Eitan Gurel, 1986, "Price and Volume Effects Associated with Changes in the S&P 500 List: New Evidence for the Existence of Price Pressures", *The Journal of Finance* **41**.4, pp. 815–829.

Hseu, Mei-Maun, Huimin Chung, and Erh-Yin Sun, 2007, "Price Discovery across the Stock Index Futures and the ETF Markets:: Intra-Day Evidence from the S&P 500, Nasdaq-100 and DJIA Indices.", *Review of Pacific Basin Financial Markets & Policies* **10**.2, pp. 215–236.

Huang, Jennifer C. and Ilan Guedj, 2009, "Are ETFs Replacing Index Mutual Funds?", *SSRN eLibrary*.

Investment Company Fact Book, 2009, Investment Company Institute.

Jank, Stephan, 2012, "Mutual fund flows, expected returns, and the real economy", *Journal of Banking & Finance* **36**.11, pp. 3060–3070.

Kacperczyk, Marcin and Amit Seru, 2007, "Fund Manager Use of Public Information: New Evidence on Managerial Skills", *The Journal of Finance* **62**.2, pp. 485–528.

Kalaycıoğlu, Serdar, 2004, "Exchange Traded Fund Flows", *SSRN eLibrary*.

Kyle, Albert S., 1985, "Continuous Auctions and Insider Trading", *Econometrica* **53**.6, pp. 1315–1335.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1992, "The impact of institutional trading on stock prices", *Journal of Financial Economics* **32**.1, pp. 23–43.

Lo, Andrew W. and Jiang Wang, 2009, "Stock Market Trading Volume", *Handbook of Financial Econometrics: Applications*.

Lu, Lei, Jun Wang, and Ge Zhang, 2009, "Long Term Performance of Leveraged ETFs", *SSRN eLibrary*.

Milonas, N. T and G. G Rompotis, 2010, "Dual Offerings of ETFs on the Same Stock Index: US vs. Swiss ETFs", *The Journal of Alternative Investments* **12**.4, pp. 97–113.

Newey, Whitney K. and Kenneth D. West, 1987, "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica* **55**.3, pp. 703–708.

Oh, Natalie Y. and Jerry T. Parwada, 2007, "Relations between mutual fund flows and stock market returns in Korea", *Journal of International Financial Markets, Institutions and Money* **17**.2, pp. 140–151.

Pearson, Neil, Allen Poteshman, and Joshua White, 2007, "Does Option Trading Have a Pervasive Impact on Underlying Stock Prices?", *AFA 2008 New Orleans Meetings Paper*.

Petajisto, Antti, 2011, "Inefficiencies in the Pricing of Exchange-Traded Funds", *SSRN eLibrary.*

Petajisto, Antti, 2009, "Why Do Demand Curves for Stocks Slope Down?", *Journal of Financial and Quantitative Analysis* **44**.05, p. 1013.

Petersen, Mitchell A., 2009, "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches", *Rev. Financ. Stud.* **22**.1, pp. 435–480.

Poterba, James M. and John B. Shoven, 2002, "Exchange-Traded Funds: A New Investment Option for Taxable Investors", *The American Economic Review* **92**.2, pp. 422–427.

Quinn, Michael, Peter Tufano, and Ryan Taliaferro, 2006, "Live Prices and Stale Quantities: T+1 Accounting and Mutual Fund Mispricing", *SSRN eLibrary.*

Rakowski, David and Xiaoxin Wang, 2009, "The dynamics of short-term mutual fund flows and returns: A time-series and cross-sectional investigation", *Journal of Banking & Finance* **33**.11, pp. 2102–2109.

Richie, Nivine and Jeff Madura, 2007, "Impact of the QQQ on liquidity and risk of the underlying stocks", *The Quarterly Review of Economics and Finance* **47**.3, pp. 411–421.

Roll, Richard, Eduardo Schwartz, and Avanidhar Subrahmanyam, 2010, "O/S: The relative trading activity in options and stock", *Journal of Financial Economics* **96**.1, pp. 1–17.

Roll, Richard, Eduardo Schwartz, and Avanidhar Subrahmanyam, 2012, "Trading Activity in the Equity Market and Its Contingent Claims: An Empirical Investigation", *SSRN eLibrary.*

Schlusche, Bernd, 2009, "Price Formation in Spot and Futures Markets: Exchange Traded Funds Vs. Index Futures", *SSRN eLibrary.*

Scholes, Myron S., 1972, "The Market for Securities: Substitution versus Price Pressure and the Effects of Information on Share Prices", *The Journal of Business* **45**.2, pp. 179–211.

Schwert, G. William, 1990, "Stock Market Volatility", *Financial Analysts Journal* **46**.3, pp. 23–34.

Shin, Sangheon and Gökçe Soydemir, 2010, "Exchange-traded funds, persistence in tracking errors and information dissemination", *Journal of Multinational Financial Management* **20**.4-5, pp. 214–234.

Shleifer, Andrei, 1986, "Do Demand Curves for Stocks Slope Down?", *The Journal of Finance* **41**.3, pp. 579–590.

Subrahmanyam, A, 1991, "A theory of trading in stock index futures", *Review of Financial Studies* **4**.1, pp. 17–51.

Subrahmanyam, A, 2008, "Lagged order flows and returns: A longer-term perspective", *The Quarterly Review of Economics and Finance* **48**.3, pp. 623–640.

Subrahmanyam, A, 2009, "The implications of liquidity and order flows for neoclassical finance*", *Pacific-Basin Finance Journal* **17**.5, pp. 527–532.

Svetina, M. and S. Wahal, 2008, "Exchange Traded Funds: Performance and Competition", *Arizona State University, Working Paper.*

Tsay, Ruey S., 2010, *Analysis of Financial Time Series*, 3rd ed., John Wiley & Sons.

Tse, Yiuman, Paramita Bandyopadhyay, and Yang-Pin Shen, 2006, "Intraday Price Discovery in the DJIA Index Markets", *Journal of Business Finance & Accounting* **33**.9-10, pp. 1572–1585.

Warther, Vincent A., 1995, "Aggregate mutual fund flows and security returns", *Journal of Financial Economics* **39**.2-3, pp. 209–235.

Warther, Vincent A., 2002, "Aggregate mutual fund flows and security returns", *Journal of Financial Economics* **39**.2-3, pp. 209–235.

Watson, John and J. Wickramanayake, 2012, "The relationship between aggregate managed fund flows and share market returns in Australia", *Journal of International Financial Markets, Institutions and Money* **22**.3, pp. 451–472.

Wermers, Russ, 1999, "Mutual Fund Herding and the Impact on Stock Prices", *The Journal of Finance* **54**.2, pp. 581–622.

Wooldridge, Jeffrey M., 2001, *Econometric Analysis of Cross Section and Panel Data*, 1st ed., The MIT Press.

A Appendix

A-1 Tables

Table A-1. Investment Advisors and their largest ETFs

The table shows investment advisors and the largest ETFs they manage. Fund total net assets (TNA) are in $millions. Number of funds under management is denoted as No of funds.

Fund Investment Advisor	Fund Ticker	Fund TNA	Fund Name	Starting date of the first ETF	No of Funds
SSGA Funds Management Inc	SPY	67,596	SPDR S&P 500 ETF TRUST	21-Jan-93	28
BlackRock Fund Advisors	IVV	21,589	ISHARES S&P 500 INDEX FUND	25-Apr-00	77
Invesco PowerShares Capital Mgmt	QQQQ	17,676	POWERSHARES QQQ NASDAQ 100	4-Mar-99	56
Vanguard Group Inc/The	VTI	13,805	VANGUARD TOTAL STOCK MKT ETF	19-Feb-93	26
Bank of New York Mellon/The	MDY	8,308	SPDR S&P MIDCAP 400 ETF TRUS	26-Apr-95	1
ProShares Advisors LLC	SDS	3,715	PROSHARES ULTRASHORT S&P500	20-Jun-06	57
Vanguard Group	VIG	2,970	VANGUARD DIVIDEND APPREC ETF	24-Apr-06	1
Rydex Investments	RSP	1,833	RYDEX S&P EQUAL WEIGHT ETF	28-Apr-03	17
Direxion Shares ETF Trust	FAS	1,706	DIREXION DAILY FIN BULL 3X	3-Nov-08	13
DB Commodity Services LLC	UUP	930	POWERSHARES DB US DOL IND BU	15-Feb-07	2
First Trust Advisors LP	FCG	382	FIRST TRUST ISE-REV NAT GAS	10-Jun-03	29
WisdomTree ETFs/USA	DLN	378	WISDOMTREE L/C DIVIDEND FUND	22-Aug-94	10
Charles Schwab Investment Mgmt	SCHX	258	SCHWAB US LARGE-CAP ETF	3-Nov-09	5
Claymore Advisors LLC/USA	CGW	201	CLAYMORE S&P GLBL WAT IDX ET	19-May-06	7
WisdomTree Investments Inc	DES	174	WISDOMTREE SMALLCAP DVD FUND	15-Jun-06	1
Global X Management Co LLC	SIL	51	GLOBAL X SILVER MINERS ETF	19-Apr-10	1
Van Eck Associates Corp	EVX	26	MARKET VECTORS ENV SERV ETF	11-Oct-06	1
WisdomTree Asset Management	ROI	19	WISDOMTREE LARGECAP GROWTH	4-Dec-08	1
Total		141,617			333