2019

# ESSAYS ON INVESTMENTS

Michael Farrell
*University of Kentucky*, mike@mikefarrell.com
Digital Object Identifier: https://doi.org/10.13023/etd.2019.149

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Michael Farrell, Student

Dr. Kristine Hankins, Major Professor

Dr. Kenneth Troske, Director of Graduate Studies

# ESSAYS ON INVESTMENTS

DISSERTATION

A dissertation submitted in partial
fulfillment of the requirements for
the degree of Doctor of Philosophy
in the Gatton College of Business
and Economics at the
University of Kentucky

By
Michael Farrell
Lexington, KY

Director: Dr. Kristine Hankins, Professor of Finance

2019

ABSTRACT OF DISSERTATION

## ESSAYS ON INVESTMENTS

The first chapter studies mutual funds. I model intraquarter trading and use a genetic algorithm to estimate the trade pattern that is most consistent with the fund's daily reported returns. I validate the model empirically on a sample of institutional trades from Ancerno and I confirm that the method more accurately predicts daily holdings when compared to existing naïve assumptions. Further, my method is substantially more accurate in classifying a fund's tendency to supply liquidity, and this increased precision has important implications for identifying superior performing funds. Specifically, a long-short strategy based on the model's liquidity provision measures earns significant abnormal returns, while a similar strategy that relies on quarterly holdings does not exhibit any outperformance. The second chapter studies investment research. We find evidence that crowdsourced investment research facilitates informed trading by retail investors and improves firm liquidity. Specifically, retail order imbalances are strongly correlated with the sentiment of Seeking Alpha articles, and the ability of retail order imbalances to predict returns is roughly twice as large on research article days. In addition, firms with exogenous reductions in Seeking Alpha coverage experience increases in bid-ask spreads and price impact, with the effect being stronger for firms with high retail ownership. Our findings suggest that technological innovations have helped democratize access to investment research with important implications for firm liquidity.

KEYWORDS: Investments, Liquidity, Performance, Mutual Funds, Crowdsourced Research

Author's signature: Michael Farrell

Date: April 30, 2019

ESSAYS ON INVESTMENTS

By
Michael Farrell

Director of Dissertation: Kristine Hankins

Director of Graduate Studies: Kenneth Troske

Date: April 30, 2019

# ACKNOWLEDGMENTS

I would like to thank my dissertation committee: Kristine Hankins, Russell Jame, Leonce Bargeron, Mark Liu, Frank Scott, and Min Young Lee for advice and guidance on my research. I would like to especially thank Russell Jame for the many hours spent advising me on this dissertation's first chapter and coauthored work in the second chapter. In addition, I thank Clifton Green, Russell Jame, and Stan Markov for their collaboration in the second chapter. I am also grateful to the entire finance faculty at the University of Kentucky: each member has been generous with their time and insights.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

**Chapter 1 Read Between the Filings: Daily Mutual Fund Holdings and Liquidity Provision**

## 1.1 Introduction

Mutual funds are required to report their equity holdings at a quarterly interval. These filings have provided both investors and researchers with an invaluable source of information. However, due to the quarterly frequency, it is not possible to know the exact timing of their trades. As a result, it remains a challenge to address many important questions. For example, does the fund tend to demand or supply liquidity? Does the fund make informed trades over short-horizons or around major information events (e.g., earnings announcements)? Does the fund engage in window dressing or portfolio pumping at the end of each quarter?

In order to study intra-quarter activity, many researchers have either made admittedly naïve assumptions about the timing of trades based on filings, or have abstracted from the stock level in favor of the fund's observable daily returns. The main advantage at the stock level is that analysis reflects actual trades that are observed in the quarterly filings; this creates a trade-off in the lack of precision when measuring trade timing and adds noise to any derived style measures since market conditions change throughout the quarter. Analysis at the level of fund returns has the potential to capture additional features of trading style related to timing, but suffers from an imprecise measure of actual trading.

Researchers have also made use of alternative datasets, such as Ancerno Ltd., in order to measure precise actions of funds. While addressing the timing of intraquarter trades, these data have several drawbacks relating to their limited coverage of funds as well as the restriction of analysis to the fund family instead of the individual mutual funds. Similarly, studies such as Campbell, Ramadorai, and Schwartz (2009) and Ha and Hu (2017) have used Trade and Quote (TAQ) data to infer daily institutional trading. Such approaches captures time-sensitive characteristics of institutional trades in the overall market but are unable to isolate the measurement to any given fund.

I use both the fund's holdings and daily returns data to develop a holdings measure for the days in between quarterly filings. The main assumption of the method is that the fund's daily returns must reflect the returns of its underlying holdings. Since all non-roundtrip equity trades are observable at the quarterly level, I estimate their precise timing by modeling and comparing a large number of possible trade sequences and select the pattern that is most consistent with the fund's daily returns. I rely on the fact that stocks trade in discrete quantities (i.e. stocks trade in units and not at infinitesimal increments). I assume that the portfolio allocation at the start of the quarter changes toward that of quarter end in discrete steps. Specifically, I assume that the manager breaks up each quarterly trade into a finite number of pieces and trades each piece on a distinct day. This creates a large combinatorial problem where many possible trade combinations exist. To compare combinations, I reconstruct the daily returns of the fund's portfolio under the hypothetical trade pattern. Absent significant inflows, purchases must be funded by sales. I therefore constrain the fund's implied daily flows from trading in order to eliminate unfeasible trade patterns. Using the genetic algorithm, I select the trade sequence that most closely replicates the fund's daily returns under the model's constraints.

As an example, consider a hypothetical portfolio invested entirely in stock A at the start of the quarter, and entirely in stock B by the end. Assuming that no other stocks are transacted, the portfolio return on any given day is the sum of the returns to each stock multiplied by their respective weights: $R_{p,t} = w_{A,t} \times R_{A,t} + w_{B,t} \times R_{B,t}$. While there are few legal restrictions that prevent a manager from buying and selling back and forth between stocks A and B, this is largely limited by trading costs. I therefore assume that each holding changes monotonically within the quarter.[1] If the initial holdings in A were 100 shares, and the ultimate holdings in B were 40 shares, I could break the sale of A into 100 pieces and the purchase of B into 40; the pieces can trade simultaneously. I could label each share in the portfolio: $A_1, A_2, ..., A_{100}, B_1, B_2, ..., B_{40}$. If there are 60 trading days in the portfolio, there are 60 possible days when stock $A_1$ could be sold. Considering stocks $A_1$ and $A_2$, there are $60^2$ possible combinations. For the entire portfolio over the quarter, there are $60^{140}$ possible combinations. Of course, many of these combinations are redundant since $A_1$ and $A_2$ are identical. Given the high computational demands that this presents, I simplify the problem by assuming that each quarterly trade is split into only four equal pieces: each piece of A is 25 shares and each piece of B is 10 shares. This reduces the number of possible combinations in this example to $60^8$. Since the trade sequences create different portfolios through time, the sequences imply different portfolio returns on each day. I assume that the trade sequence that best reproduces the fund's daily returns is the most likely representation of the fund's actual trade pattern. To identify this sequence, I use the genetic algorithm to efficiently search the set of hypothetical trade sequences and select the one with the closest fit on the fund's daily returns.[2]

I validate the method empirically using a sample of institutional trades from ANcerno Ltd. Specifically, I infer holdings at quarterly intervals and apply the method to compare the observable daily fund holdings with those estimated from the model. Compared to the the standard assumption of end-of-quarter trades, the method reduces the mean squared error (MSE) of dollar-weighted daily holdings by 73%.

As an application, I use the model to estimate the fund's tendency to supply liquidity. In this context, liquidity provision is the trading against short-term mispricing from non-fundamental pressure (Nagel, 2012). The liquidity supplier provides immediacy at favorable prices and realizes a profit as prices later reverse. Given the

[1]Given the low dimensionality of this example, the daily trades could be solved analytically through a system of equations. For large portfolios, the system would eventually become underdetermined and the method would fail. By forcing trades to take place in discrete quantities, a solution can be found.

[2]Given that the four pieces of a given quarterly trade are identical, there are several sequences that will produce identical holdings and return patterns by symmetry. I accept the pattern selected by the algorithm, which is an approximation and involves a stochastic process.

importance of timing, this trading style is difficult to estimate with low frequency data. Jame (2018) develops a measure of the fund's tendency to supply liquidity based on short-term momentum trading and finds that this predicts future returns among hedge funds. Using observable daily trades, I replicate the momentum measure of Jame (2018) within the Ancerno sample and examine how well it can be approximated using the algorithm-estimated trades. Compared to the benchmarks of quarter-end trading and uniform trading across the quarter, the algorithm-estimated trades yield a significantly stronger approximation of the liquidity provision measure.

After establishing validity within Ancerno, I extend my analysis to the broad sample of mutual funds. I estimate the holdings in between filings, which are unobservable given current disclosure requirements. I then replicate the trade-based liquidity provision measure of Jame (2018) using the estimated intra-quarter trades. The main finding is that liquidity provision is a strong predictor of future fund performance. A portfolio long funds in the bottom quintile and short funds in the top quintile of momentum over the past two quarters produces an annualized abnormal return of 1.9%. This finding is robust to the inclusion of the traded liquidity provision factor of Rinne and Suominen (2014), indicating that the measure provides incremental information in the classification of the fund's liquidity demand.

This paper makes two contributions to the finance literature. First, I develop a measure of daily holdings using the quarterly holdings and daily returns data that is widely available to academic researchers. I show that this is a more accurate approximation of the fund's intraquarter trading than existing alternatives. My method can be applied towards additional topics that require shorter frequency trade data among mutual funds. Second, by estimating trades at a daily level, I measure the fund's tendency to trade against short term price reversals. I show that this measure predicts future fund performance. This compliments a growing literature that links liquidity provision with fund performance (Franzoni and Plazzi, 2013; Jame, 2018; Rinne and Suominen, 2014). Finally, I show that the relation between liquidity provision and future returns is not anticipated by investors in the form of flows.

The rest of the paper is organized as follows: Section 2 outlines the method for estimating intraquarter holdings and trading. In Section 3, I validate the model using the Ancerno sample of institutional trades. Section 4 applies the methodology to available sample of US active mutual funds. Section 5 concludes.

## 1.2 Model

### 1.2.1 Intraquarter Holdings

Traditional methods of analyzing mutual fund trading have relied on measures derived from holdings and returns data separately. Holdings measures directly reflect portfolio holdings and are better at capturing investment styles (Daniel, Grinblatt, Titman, and Wermers, 1997). On the other hand, meaningful activity takes place in between filings, producing a persistent "return gap" Kacperczyk, Sialm, and Zheng (2006). To examine additional aspects of trading, albeit without the precision of holdings data, many studies have relied solely on returns data (Carhart, 1997; Fama and French, 2010).

In this section, I develop an approach to estimate the fund's daily holdings using daily returns data for both the underlying holdings and the fund itself. As a starting point, Kacperczyk et al. (2006) approach these data jointly in their "return gap" measure. Specifically, they consider the returns to a hypothetical portfolio that holds the fund's initial holdings without trading and compare its returns to the those of the actual fund. My method attempts to reproduce the fund's daily returns by selecting the timing of the trades that are observed in quarterly filings. Since the fund's daily returns must reflect their underlying holdings, the trading pattern that best reproduces its returns is likely an informative estimate of daily holdings.

### 1.2.2 Model Assumptions

If the fund's holdings were known throughout the quarter, its daily returns could be closely inferred from the underlying security prices. Instead, the holdings are revealed only at quarter-end and the interim portfolio composition is unobservable. Depending on one's assumptions, there could exist an infinite set of possible trade sequences available to a fund manager that would, ex-post, produce identical portfolios observed in the quarterly filings. With zero trading costs and fractional share holdings, it would be possible to take on any imaginable portfolio allocation at any point within the quarter and then fall upon a particular allocation at the time of disclosure. This has posed a challenge for the study of mutual funds, resulting in a "black-box" approach to intraquarter holdings.

In order to create a tractable problem, I put forward three assumptions of intraquarter trading that bound the relevant set of possible actions available to the manager: (1) shares trade in discrete quantities, (2) the incremental effect of round-trip trades is uncorrelated with portfolio returns, and (3) all trades take place at the day's closing price. Assumption (1) is true since shares are typically traded in units greater than 1. Assumptions (2) and (3) may not be absolutely true, but serve as reasonable approximations. For instance, the majority of institutional trading takes place at the end of the trading day (Corrie Driebusch, 2018); for this reason I assume all trades are executed at the day's closing price. Of course, the degree to which these assumptions deviate from reality will impact the accuracy of the model's predictions. This is an empirical question, which I address later in this paper.

Assumptions (1) and (2) dramatically reduce the possible holdings that the fund can take on within the quarter. Given discrete trades and by neglecting round trip trades, the holdings are assumed to move monotonically in discrete steps over the quarter. Assumption (3) specifies the execution prices and ties the daily holdings to the daily returns of the fund. Taken together, the three assumptions reduce the problem to a finite set of possible trades sequences that can be compared objectively to one another based on how well they reproduce the fund's daily returns.

To formalize this problem, I consider the returns to the fund's portfolio on any given day:

$$R_{t,F} = \sum_{i=1}^{N} R_{i,t} \times w_{i,t}, \tag{1.1}$$

where $R_{t,F}$ is the fund return on day t; $R_{i,t}$ and $w_{i,t}$ are the returns and weights to each stock i on day t. Since I assume that all trades takes place at the end of the day, the portfolio weights are determined by the holdings on the previous day:

$$w_{it} = \frac{S_{i,t-1} \times P_{i,t-1}}{\sum_{i=1}^{N} S_{i,t-1} \times P_{i,t-1}}, \tag{1.2}$$

where $S_{i,t-1}$ and $P_{i,t-1}$ represent the number of shares held and the closing price of a given stock on day t-1. On any given day, $S_{i,t}$ is equal to the number of shares held at the start of the quarter, plus the cumulative share purchases from days 0 to t:

$$S_{i,t} = S_{i,0} + \sum_{d=1}^{t} \Delta S_{i,d} \tag{1.3}$$

Combining equations (1.1), (1.2), and (1.3), the fund's return on a given day can be approximated by:

$$R_{t,F} \approx \sum_{i=1}^{N} R_{i,t} \times \frac{\left(S_{i,0} + \sum_{d=1}^{t-1} \Delta S_{i,d}\right) \times P_{i,t-1}}{\sum_{i=1}^{N} \left(S_{i,0} + \sum_{d=1}^{t-1} \Delta S_{i,d} \times P_{i,t-1}\right)} = R_{t,M}. \tag{1.4}$$

$R_{t,M}$ is the return to the modeled portfolio on day t. I rely on my prior assumptions in order to put restrictions on $\Delta S_{i,d}$. First, I restrict the number of shares traded to be integers less than or equal to the total number of shares traded within the quarter, $\Delta S_i$:

$$\Delta S_{i,d} \in \{1, 2, 3, \ldots, \Delta S_i\}, \tag{1.5}$$

and

$$\Delta S_i = S_{i,T} - S_{i,0}, \tag{1.6}$$

where subscripts 0 and T designate the start and end dates of the quarterly holding period. Next, for each stock, the total of daily trades must equal the difference in trades observed in the filings:

$$\Delta S_i = \sum_{d=1}^{T} \Delta S_{i,d} \tag{1.7}$$

The assumption that round-trip trades can be ignored implies that the modeled trades will move the holdings of a given stock monotonically throughout the quarter. This further implies that for each stock, the absolute value of the total shares traded equals the sum of the absolute value of each daily trade:

$$|\Delta S_i| = \sum_{d=1}^{T} |\Delta S_{i,d}|. \tag{1.8}$$

The objective is to select the trade sequence most consistent with the available data and the model assumptions. I formulate an optimization problem where the difference between the modeled and the true fund returns are minimized throughout the quarter. Specifically, I minimize the sum of the squared deviations of the modeled and true portfolios by selecting different values of each $\Delta S_{i,d}$:

$$\min_{S_{1,1}\dots S_{N,T}} \sum_{t=1}^{T} \left( R_{t,F} - \sum_{i=1}^{N} R_{i,t} \times \frac{\left(S_{i,0} + \sum_{d=1}^{t-1} \Delta S_{i,d}\right) \times P_{i,t-1}}{\sum_{i=1}^{N} \left(S_{i,0} + \sum_{d=1}^{t-1} \Delta S_{i,d}\right) \times P_{i,t-1}} \right)^2 \tag{1.9}$$

subject to the constraints described in equations (1.5) and (1.8).

An analytical approach to this minimization problem is not possible: the objective function itself does not have a unique minimum without simultaneously considering the two constraints. Instead, this can be approached as a combinatorial optimization problem: I select the best sequence among the finite set of possible alternatives.

### 1.2.3 Selecting the Optimal Trade Sequence

The minimization problem requires the selection of the trading sequence that minimizes the sum of squared differences between the modeled and the true portfolio's daily returns. Given the constraints, there exists a finite set of sequences, however this set is very large. Consider an additional simplification that the fund trades each stock only once throughout the quarter. This effectively forces all but one $\Delta S_{i,d}$ for each stock equal to zero, with a single value of $\Delta S_{i,d}$ set to $\Delta S_i$. This reduces the dimensionality of the problem substantially. Instead of having $N \times T$ decision variables, we are left with $N$ choices: one for each stock in the portfolio. Furthermore, we can reformulate the choice variable, not as the value of $\Delta S_{i,d}$ on any given day, but instead select which day in the quarter, $k_i$, the holdings of a given stock move from $S_{i,0}$ to $S_{i,T}$. With this assumption, I redefine the value of each $S_{i,k_i,t}$ as a function of $S_0$, $S_1$, $t$, and choice variable $k_i$:

$$S_{i,k,t} = \begin{cases} S_{i,0} & t < k_i \\ S_{i,T} & t \geq k_i \end{cases}$$

$$k_i \in \{1, 2, 3, \dots, T\} \tag{1.10}$$

This redefinition effectively embeds the constraints into the objective function, yielding a simplified problem:

$$\min_{k_1\dots k_N} \sum_{t=1}^{T} \left( R_{t,F} - \sum_{i=1}^{N} R_{i,t} \times \frac{S_{i,k_i,t-1} \times P_{i,t-1}}{\sum_{i=1}^{N} S_{i,k_i,t-1} \times P_{i,t-1}} \right)^2 \tag{1.11}$$

This combinatiorial problem remains non-trivial: for a portfolio of 75 stocks in a quarter of 60 trading days, there are more than $2 \times 10^{133}$ different possible trade sequences.[3] This is larger than the estimated number of atoms in the universe, by many orders of magnitude.

[3] There are T possibilities for each stock. The total possible sequences is $T^N$.

### Relaxing the Assumption of a Single Trade per Quarter

The assumption of a single trade per stock per quarter is likely overrestrictive, given that institutional trades are typically spread over time (Barclay and Warner, 1993). In order to allow for multiple trades on a given stock, I split the fund's holdings of each stock into four equal portions. The algorithm treats each portion of a holding as an individual stock and selects the trade date for each portion independently. For example, if a fund purchased 100 shares of a given stock in a quarter, I break the quarterly trade into four distinct trades of 25 shares. The algorithm then selects the purchase date for each bundle of 25 shares that is most consistent with the model. Once a solution is found, I recombine the individual portions of each stock to calculate the daily holdings of the stock within the fund's portfolio.

### Implied Daily Fund Flows

When combined, the fund's returns and holdings imply a daily level of fund flows, given the model's assumption that the portfolio is invested entirely in equities. To improve the selection mechanism, I impose a restriction against solutions that imply infeasible daily fund flows. Assuming that the total assets are equal to the sum of shares multiplied by their respective prices, it can be shown that the model's implied flows on day t are equal to:

$$flow_t = \frac{\sum_{i=1}^{N} S_{i,t} \times P_{i,t} - \left(\sum_{i=1}^{N} S_{i,t-1} \times P_{i,t-1}\right) \times (1 + R_{t,F})}{\sum_{i=1}^{N} S_{i,t-1} \times P_{i,t-1}}. \tag{1.12}$$

Given that daily fund flows are typically small and to allow for coarseness in the modeled trades, I add a penalty of -0.01 to the fitness function for each day that implied flows exceed 2.5%. A penalty function is a technique to add constraints when bounding the set of acceptable solutions (Yeniay, 2005). The penalty has to be large enough that the algorithm heavily disfavors solutions that violate the constraint. With the constraint, the problem becomes:

$$\min_{k_1 \ldots k_N} \sum_{t=1}^{T} \left[ \left( R_{t,F} - \sum_{i=1}^{N} R_{i,t} \times \frac{S_{i,k_i,t-1} \times P_{i,t-1}}{\sum_{i=1}^{N} S_{i,k_i,t-1} \times P_{i,t-1}} \right)^2 + 0.01 V_t \right], \tag{1.13}$$

where $V_t$ is an indicator variable equal to 1 if the implied flows constraint is violated and zero otherwise. I minimize the function in (13) by selecting values of $(k_1, \ldots, k_N)$ according to the genetic algorithm.

### 1.2.4 The Genetic Algorithm

The genetic algorithm is a biologically inspired procedure used to optimize complex discrete mathematical problems. The link to biology is an analogy that compares the decision variables to a chain of DNA with genetic information. The algorithm selects the optimal solution over numerous generations based upon the principles of

Darwinian evolution: the fit reproduce while the remaining candidates are discarded (McCall, 2004). In the context of a mathematical problem, the fitness of a candidate solution is equal to the evaluation of the objective function. I transform the minimization into a maximization problem by multiplying the objective function by negative 1.

For each fund and quarter, I begin with a population of randomly generated candidate solutions. For each candidate, I evaluate its fitness based on equation (1.13). Each candidate solution is then ranked according to its fitness. Those with the highest fitness are selected with greater probability for replication. Within the replication group, solutions are paired at random to create child solutions for the next generation. For each solution variable, the child takes on a value selected from either parent at random, with a small number of values perturbed by a random mutation. This procedure is repeated for a large number of generations, converging to a solution.

To implement the genetic algorithm, I select values for the hyperparameters. I use trial and error in order to optimize performance along computational speed and convergence to a solution, however, I do not measure the quality of the solutions in the selection of hyperparameters.[4] I generate an initial population of 7,500 random candidate solutions for the first generation. After computing the fitness of each candidate, I select the top 1,200 with an additional 300 solutions at random to form the replication group of 1,500. From this group, each randomly matched solution pair creates 10 child solutions. This results in a stable population size of 7,500 each generation. Within each candidate solution, each choice variable ($k_i$), has a 1% chance of mutation, where the value is changed to a random integer between 1 and T. After repeating the procedure for 300 generations, I select the solution with the highest fitness.

## 1.3 Model Validation with Trade Level Data

The model relies on several assumptions designed to approximate the fund's daily trading. In this section, I examine the validity of the model empirically using a sample of institutional equity trades from ANcerno Ltd. (formerly the Abel Noser Corporation). The sample contains detailed trade information including the transaction price, the direction of the trade, and the day on which the trade took place. This offers an ideal setting to validate the model since the typically unobservable data are disclosed. ANcerno records trades at the manager level on behalf of their specific clients and this data became availabe to researchers for years 1999-2011. I consider

[4] The genetic algorithm is a stochastic optimization technique, which causes final solutions to vary from one another even with the same initial conditions. Solutions in this context are in fact only approximations. Selecting a large population slows down the total computation speed, however a small population may not contain sufficient information to converge to the optimal solution. Techniques exist for hyperparameter optimization, see Bergstra and Bengio (2012) for example. While this would improve the model fit, it would require additional cross-validation tests. Further, it is not guaranteed that the hyperparameters would be optimal for use on the broader mutual fund data (S12 and daily returns).

the manager-client pair as a unique fund for the purpose of the analysis.[5]

Similar to Jame (2018), Puckett and Yan (2011) and other studies that use ANcerno data, I aggregate each fund's trades over its first 36 months in the sample, creating a measure of daily holdings. In order to avoid funds that do not trade frequently, each fund must make at least one trade each month to remain in the sample. I then remove positions less than zero and merge with the CRSP daily securities file based on the CUSIP and day for each stock. I compute daily returns for the fund based on the daily holdings, daily trades, execution prices, and the daily closing price of each stock. This yields a sample of daily holdings and returns for institutional funds trading US equities. For quarterly holdings measures, I record the holdings as of the quarter end dates: December 31, March 31, June 30, and September 30.

### 1.3.1 Empirical Validation

I use a genetic algorithm to minimize equation (1.13) for each fund-quarter in order to find the dynamic portfolio that best replicates fund's observed daily returns. For ease of interpretation, I refer to the fitness of the function as the evaluation of equation (1.13) multiplied by -1. This allows for higher values of fitness to be interpreted as a greater fit on the data. Figure 1 plots the histogram of fitness values for the sample of funds in Ancerno. Since a fitness of zero would indicate a perfect fit, the distribution is truncated on the right. The distribution is therefore negatively skewed with a mean (median) fitness value of -0.020 (-0.007).

To measure the model's goodness-of-fit on the daily holdings, I estimate each individual fund's portfolio allocation within each quarter in the ANcerno sample and compare this with its observed allocation. At the end of each day, I calculate the dollar-weighted error between the modeled and the true portfolio allocation:

$$Error_t = \sum_{i=1}^{N} \left( \frac{(S_{i,t,M} - S_{i,t,O}) \times P_{i,t}}{TNA_t} \right)^2, \tag{1.14}$$

where $S_{i,t,M}$ and $S_{i,t,O}$ represent the modeled and the observed number of shares of stock $i$ held by the fund on day $t$. $TNA_t$ is the total net assets of the fund on day $t$. I calculate total net assets as the total value of all equity holdings at the end of day $t$. To measure the overall fit, I calculate the mean square error (MSE) of the dollar weighted allocation by taking an equal weighted average of each $Error_t$ throughout the quarter:

$$MSE = \frac{1}{T} \sum_{t=1}^{T} Error_t. \tag{1.15}$$

[5]The funds in this sample are actual institutional investors, however all trades need not be executed by mutual funds. For example, several studies Jame (2018); Puckett and Yan (2011); Cohen, Lou, and Malloy (2016) have used Ancerno outside of the mutual fund context.

I compare the model with alternative assumptions of intraquarter holdings. The path of each $S_{i,t,T}$ throughout the quarter is nonlinear, which invalidates linear regressions and the commonly used R-squared as a metric for goodness-of-fit (Spiess and Neumeyer, 2010). To evaluate the model's performance, I compare its MSE with that of four alternative assumptions: trades that take place at the beginning, middle, end, or that take place evenly throughout the quarter.

Table 1.1 presents the average MSE for the model and the alternative assumptions. Directly comparing the average MSE among alternatives is uninformative, given the skewness in both the fitness and the MSE distributions. In column (2), I calculate the average ratio of the model's MSE ($MSE_{model}$) to the alternative. The average ratio ranges from 0.23 for the assumption that trades take place at the end-of-quarter to 0.79 for the portfolio that trades evenly throughout the quarter. To verify that the ratios are statistically less than 1, I apply a t-test on the log of the ratio. In each case, the log-transformed value is less than zero, statistically significant at the 1% level. Overall, this indicates that the modeled portfolio holdings approximate the true portfolio's intra-quarter holdings better than the existing alternatives.

### Regressions: Model Fitness and Holdings MSE

The central assumption of the model is that the trade pattern that best replicates the fund's daily returns also contains information about the true underlying holdings within the quarter. To confirm that the fit of the model in equation (1.13) corresponds with a stronger prediction of holdings, I investigate the relation between fitness and the quarter's holdings MSE. Intuitively, fund-quarters with a greater algorithm fitness should have a better fit on the daily holdings. To test this, I regress the MSE of the holdings onto the fitness of the modeled portfolio:

$$ln(MSE_{holdings}) = \alpha + \beta \times ln(fitness_{model}) + \gamma' Controls + \epsilon. \tag{1.16}$$

To account for mechanical relations explained by fund characteristics, I include controls for the number of stocks in the portfolio, the total net assets of the fund, the ratio of stocks traded to total stocks in the portfolio, as well as fund and quarter fixed effects. All continuous variables are log transformed in order to measure elasticities.[6] Given the potential of residual correlation across fund and time (Petersen, 2009), standard errors are clustered at the fund and quarter. Table 1.2 presents the regression results of equation (1.16). A 1% increase in fitness coincides with an estimated reduction in the MSE of approximately 0.4%. Overall, this indicates that the strength of the model's fitness corresponds with its ability to predict holdings in between quarterly filings.

Next, I verify that the fitness of the algorithm estimate corresponds with an improved measure of intraquarter holdings *relative* to alternative assumptions. In Panel B, I replace model MSE on the left hand side of equation (1.16) with the ratio of the model MSE over the MSE from the next best alternative, the evenly traded portfolio. The coefficient $\beta$ can be interpreted as the marginal improvement in the

[6] Since the domain of fitness is $(-\infty, 0]$, I use the following transformation: $-log(-fitness)$.

estimate of intraquarter holdings relative to the. A 1% increase in fitness coincides with an estimated reduction in the MSE ratio by approximately 0.08%.

### Liquidity Provision in the Ancerno Sample

As an addition test, I verify that the modelled trades indicate a similar liquidity provision trading style when compared with the actual trades in ANcerno. In this context, liquidity provision is the trading against short-term mispricing from non-fundamental pressure (Nagel, 2012). The liquidity supplier provides immediacy at favorable prices and realizes a return as prices later reverse.

To estimate the fund's tendency to supply liquidity, I use the trade-based momentum measure of Jame (2018). For each stock on each day, the previous one-day and five-day market-adjusted returns are calculated. The momentum measure, Mom1&5, is defined as the average of the 1 and 5 days market-adjusted returns. Since non-fundamental price movement causes reversals, liquidity supplying trades are defined as purchases (sales) in stocks that have decreased (increased) in price over the recent past. At the fund level, the momentum measure is calculated quarterly by taking the dollar-volume weighted average Mom1&5 of stocks purchased minus the dollar-volume weighted average of stocks sold. Funds with higher (lower) levels of momentum trading are defined as liquidity demanders (suppliers) since they suffer (benefit) from the price subsequent reversals in their trades.

Table 1.3 presents the correlation matrix between the momentum measure estimated on the observed daily trades (Mom1&5), the algorithm ($Mom1\&5_{calc}$), the evenly traded ($Mom1\&5_{smooth}$), and the quarterly momentum measure of Jame (2018) (Mom1&5Q). The last two measures are almost identical as they both assume trading is equally likely within the quarter.[7] Among the approximations of the momentum measure from observed daily trades (Mom1&5), the algorithm-calculated measure has the highest correlation value of 0.76. This suggests that the daily estimated trades are informative for estimating trading style.

Next, I compare the measures in a multivariate setting. Since Mom1&5Q and $Mom1\&5_{smooth}$ are almost perfectly correlated ($r = 0.98$), I omit $Mom1\&5_{smooth}$. Standardizing the variables to be mean zero and standard deviation of one, I regress the momentum measure from observed daily trades (Mom1&5) onto the momentum measure from the estimated daily trades ($Mom1\&5_{calc}$) and the quarterly momentum measure (Mom1&5Q). Table 1.3, Panel B presents the results. Both coefficients are positive and statistically significant; however, the algorithm calculated measure is stronger with a coefficient of 0.51 compared to 0.19 for Mom1&5Q. An F-test rejects the null hypothesis that the coefficients are equal. This suggests that the estimated daily trades provide a stronger approximation of the derived momentum trading style measure than alternative existing assumptions.

[7] Mom1&5Q is calculated as if the trade takes place on the last day of the quarter and evenly weighs Mom1&5 throughout the quarter. This produces a slightly different measure from the evenly traded assumption. The evenly traded pattern is calculated as if the trades take place evenly throughout the quarter and Mom1&5 is averaged accordingly. The difference between the two measures come from intraquarter variation in stock prices and portfolio weights.

## 1.4 Application: US Active Mutual Funds 2002-2016

In the previous section, I used a sample of observable institutional trades to demonstrate that the model provides an improved estimate of both intraquarter holdings and a derived measure of liquidity provision style. In this section, I apply the methodology to the sample of US active mutual funds from 2002-2016, in line with when funds consistently began reporting their holdings quarterly.[8] The precise estimate of intraquarter holdings enables a trade-based measure of liquidity provision previously unavailable to mutual funds. I use this measure to show the predictive relation between fund liquidity provision on future returns.

### 1.4.1 Mutual Fund Data and Methodology

Mutual fund holdings data come from the Thomson Reuters S12 file. Daily fund returns and NAV data, as well as quarterly reported fund characteristics are drawn from the Center for Research in Security Prices (CRSP). I aggregate returns and holdings to the fund level and use MFLINKS to merge the two databases. Following Jordan and Riley (2016), I apply several screens to ensure that the funds are actively managed and invested primarily in domestic equities: funds must not be identified as index funds, ETFs, or annuities. An additional screen for terms within the fund name is applied to remove funds not primarily invested in U.S. equities. I restrict the sample to funds with a minimum of 90% in equities and a maxiumum of 10% in cash.

In order to ensure that the data is reliable, I follow Coval and Stafford (2007) and impose several restrictions: quarterly flows must be between $-50\%$ and 200% and total net assets must have been less than \$1 billion in the past. In addition, I require that the ratio of TNA between the two databases not exceed 1.1 (or fall below 1/1.1) and the fund must have at least 20 holdings.

For each fund-quarter, I estimate intraquarter holdings by finding the trade pattern that is most consistent with the fund's observed daily returns. Since returns are reported for the entire fund, including cash holdings, I divide the reported daily return by the one minus the fraction of the fund's cash holdings:

$$R_{t,F} = \frac{R_{t,F,reported}}{(1 - \%Cash)} \tag{1.17}$$

This transformation offsets the effects of the cash drag on the estimate of the fund's equity portfolio returns. Funds hold cash for a variety of reasons particularly to meet redemptions, especially when flows are volatile (Chordia, 1996; Yan, 2006). Given that cash needs can vary throughout the quarter, I add the percentage of cash holdings to the model. I replace $R_{t,F}$ in equation (1.13) with the right hand side of

[8] Prior to this, most funds reported semi-annually. Quarterly holdings disclosures were mandated only at the fund family level through the 13-F filings.

equation (1.17):

$$\min_{k_1 \ldots k_N, \%Cash} \sum_{t=1}^{T} \left[ \left( \frac{R_{t,F,reported}}{(1 - \%Cash)} - \sum_{i=1}^{N} R_{i,t} \times \frac{S_{i,k,t-1} \times P_{i,t-1}}{\sum_{i=1}^{N} S_{i,k,t-1} \times P_{i,t-1}} \right)^2 + 0.01V_t \right], \tag{1.18}$$

Since this remains an discrete problem, I restrict the estimation of the portfolio invested in cash, $\%Cash$, to integers between 1 and 10 percent. This approximation could be relaxed to allow for smaller intervals and added precision, however I must trade off precision for computational resources.[9] This reflects the method's major trade-off: I accept an approximation in order to create a problem that can be solved.

### 1.4.2 Liquidity Provision of Active Mutual Funds

In order to estimate the daily trades of US active mutual funds, I minimize equation (1.18) using the genetic algorithm. To reduce noise in the estimates, I restrict the sample to funds with a fitness value greater than -0.15.[10] Based on the trades, I repeat the analysis from section 3.1.2 and calculate the fund's momentum trading measure each quarter.

Table 1.4, Panel A presents summary statistics for the final sample of 64,480 fund-quarters. The mean $Mom1\&5$ estimates are positive indicating that, on average, mutual funds demand liquidity. This is consistent with Rinne and Suominen (2014) and the negative average value for $\beta_{RLP}$. Panel B presents the correlation matrix: $Mom1\&5_{calc}$ and $Mom1\&5Q$ are are highly correlated ($r = 0.51$), consistent with their relation estimated within the Ancerno sample ($r = 0.58$) in Table 1.3. The two measures are negatively correlated with $\beta_{RLP}$ by design: higher values of momentum are consistent with liquidity demand, whereas higher values for $\beta_{RLP}$ are consistent with a greater exposure to the priced liquidity provision factor of Rinne and Suominen (2014). Overall, this suggests that $Mom1\&5_{calc}$ shares information about the fund's tendency to supply liquidity with the existing regression-based measure, however the measures capture different aspects of this style given the modest correlation coefficient of -0.06.

To examine the persistence of the fund's tendency to supply liquidity, I sort the funds into quintiles each quarter based upon their level of $Mom1\&5_{calc}$ over the prior two quarters. If a fund follows a consistent liquidity trading style, the quintiles averages should preserve a similar ordering of $Mom1\&5_{calc}$. Figure 2 presents the average level of $Mom1\&5_{calc}$ based upon the prior two quarter quintile sorts. The monotonic increase in the average level of $Mom1\&5_{calc}$ suggests that past estimates of the fund's tendency tend to predict future liquidity demand.

[9] For example, allowing the average cash holdings to vary at intervals of 0.5% would increase the precision of the final estimates, however this would require greater resources. In addition, the selection the optimal interval is an additional optimization problem in and of itself.

[10] Fitness is defined as previously: the objective function (equation 1.18) multiplied by negative 1. Restricting funds to fitness greater than -0.15 eliminates 33,907 fund-quarters. Lowering the fitness threshold produces qualitatively similar results.

Given the measure's persistence, I investigate the relation between the momentum trading and future fund returns. In Figure 3, I compare the fund returns from quintile sorts based the previous two quarters' $Mom1\&5_{calc}$. There is a pattern associating greater liquidity provision (low $Mom1\&5_{calc}$) with higher performance: average returns decrease almost monotonically from the top quintile (3.2 bps) to the bottom quintile (-9.3 bps). Taken together, the univariate evidence in Figures (2) and (3) suggests that liquidity provision is persistent over time and this can be used to predict future returns. Next, I turn to a multivariate regression setting to investigate whether the measure can predict abnormal returns.

### 1.4.3 Liquidity Provision and Abnormal Returns

To test whether the momentum factor is associated with future abnormal returns, I use the quintile sorted portfolios based upon the prior two quarters' $Mom1\&5_{calc}$. I create a new portfolio that is long the portfolio of quintile 5 and short that of quintile 1, rebalancing every month for the length of the time series. I regress the portfolio returns $R_{q5-q1}$ onto the Fama-French 4-factor model, augmented by the $R_{LP}$ liquidity provision factor of Rinne and Suominen (2014).

Table 1.5 presents the regression results. In specification (1) , there is at -16 basis point abnormal return to the long minus short portfolio. This result indicates that high momentum funds significantly underperform on a risk-adjusted basis when compared to low momentum funds. To examine how an individual can more profitably invest with this information, I estimate the abnormal returns for the low and the high momentum portfolio separately. The low portfolio, presented in specification (2), has an alpha of -6 basis points, which is statistically indistinguishable from zero. The portfolio formed from the funds with the highest values of momentum underperform with an alpha of -16.64 basis points. This suggests that the difference in abnormal returns comes from the most liquidity demanding quintile. An investor can receive higher returns simply by avoiding the most liquidity demanding funds in terms of $Mom1\&5_{calc}$.

### 1.4.4 Liquidity Provision and Cross-sectional Fund Returns

To test whether the fund's liquidity provision style predicts future cross-sectional returns, I estimate 165 monthly Fama-Macbeth (1973) regressions. Specifically, each month I regress fund excess returns onto measures of liquidity provision with common control variables in the mutual fund literature:

$$R - R_f = \alpha + \beta \times Liquidity\ Provision + \gamma' Controls + \epsilon. \tag{1.19}$$

Standard errors are corrected for potential time-series dependence using the Newey-West (1987) procedure. Liquidity provision measures, $Mom1\&5_{calc}$, $Mom1\&5_{end}$, $Mom1\&5Q$, and $\beta_{RLP}$ are standardized to have zero mean and unit standard deviation. $Mom1\&5_{calc}$ and $Mom1\&5Q$ are averaged over the past two quarters; $\beta_{RLP}$ is estimated over the past two years using monthly returns data (Rinne and Suominen, 2014) and is rank transformed. Control variables include expenses, turnover, cash, fund age, total net assets, and percent flows. All controls are expressed as percentages, except for TNA and fund age, which are by transformed the natural logarithm. All controls are lagged by one quarter, except fund flows, which are lagged by one, two, and three months.

Table 1.6 presents the regression results. Specification (1) through (5) do not include fixed effects; specification (6) includes fund style fixed effects based upon CRSP style codes. Overall, the results are consistent with liquidity suppliers achieving higher returns in the cross-section. A one standard deviation decrease in $Mom1\&5_{calc}$ is associated with an estimated marginal increase in excess returns between 4.4 and 5.9 basis points. In specification (6), with all the control variables, the marginal effect is 5.9 basis points in absolute terms. Momentum measures based on solely quarterly data, $Mom1\&5Q$ and $Mom1\&5_{end}$, do not predict future returns with statistical significance. The coefficient on the rank transformed $\beta_{RLP}$ is positive but not statistically significant. Overall, this evidence suggests that the daily holdings-based liquidity provision measure is associated with future fund excess returns, controlling for a variety of factors. This prediction is incremental above the existing regression based measure.

### 1.4.5 Liquidity Provision and Fund Flows

I examine whether investors anticipate future fund performance from liquidity provision. Investors typically reward funds for outperforming benchmarks with greater inflows (Del Guercio and Tkac, 2002; Ivković and Weisbenner, 2009; Sirri and Tufano, 1998), however there has been mixed evidence on whether this predicts future returns (Frazzini and Lamont, 2008; Zheng, 1999). If investors recognize the fund's tendency to supply liquidity and its relation with future returns, the momentum measure could capture future fund flows. To test this relation, I run a panel regression of monthly flows onto $mom1\&5_{calc}$:

$$flow_t = \beta_0 + \beta_1 \times mom1\&5_{calc} + \gamma' Controls + \epsilon. \tag{1.20}$$

Controls follow Del Guercio and Tkac (2002) and include: an indicator for outperforming the S&P 500, lagged excess returns (separated above and below the S&P 500), Jensen's alpha (separated above and below the S&P 500), and the tracking error (separated above and below the Jensen's alpha above the lagged excess returns above the S&P 500), fund age, TNA, lagged flow, and year × style dummies. Overall these measures control for the fund flows that result from performance without an explicit recognition of liquidity provision.

Table 1.7 presents the regression results. The first column's measure of flows is in millions of dollars; the second column uses a percent flow expressed as fraction between 0 and 1. Overall, investors reward performance: lagged excess returns and Jensen's alpha are strongly associated with inflows. In addition investors react differently to diversifiable risk, measured with tracking error, depending on the state of performance. A one percent increase in tracking error is associated with a 0.5% inflow when performance is above the S&P 500, however this is associated with a 2.1% outflow when the fund underperforms the benchmark.

The variable of interest, $mom1\&5_{calc}$ , has a small estimated effect on flows and is statistically indistinguishable from zero. The 95% confidence interval for the marginal effect of a one standard deviation of $mom1\&5_{calc}$ on the percent flows is between -0.004% and 0.004%. Given how closely the confidence interval straddles zero, the relationship between liquidity provision and future fund flows is very modest if it exists at all. This suggests that investors investors tend not to respond very strongly to the fund's liquidity provision style.

### 1.5 Conclusion

This paper develops a method to estimate mutual fund positions in between mandated filings using the fund's daily returns and its quarterly holdings. Using a sample of observable institutional trading, I show that the method yields an improved estimate of both intraquarter trading and a trade-based liquidity provision measure versus existing naïve assumptions. I apply the method to the broad sample of mutual funds from 2002 to 2016. Using the estimated daily holdings, I show that the fund's tendency to supply liquidity is persistent over time and that this characteristic predicts future abnormal returns.

The main methodological contribution of this paper is the synthesis of both observable fund characteristics, individual stock level data, and known economic relationships, such as the definition of a portfolio return, into a single measure of daily holdings. This provides two clear advantages: (1) by construction, the measure contains more information than its components taken separately and (2) the measure reflects actual trades which facilitates interpretation and applicability to various problems. This is particularly useful for a trading style measure such as liquidity provision: the momentum measure can be estimated with a single quarter of data using trades that are observed at the quarterly level.

The model can be extended to consider more data that are commonly available in financial research. For example, the total trading volume of a given stock could be used in the model with all mutual fund trading estimated simultaneously over a single quarter. This for example, could bound the selection space and improve the accuracy of the predictions. A separate improvement could involve machine learning over the limited existing daily trade data in order to uncover empirical regularities in mutual fund trading.

Finally, the methodology outlined in this paper can be applied to a range of topics surrounding mutual fund behavior and performance. The estimated daily holdings could be used to measure various aspects of trading such as investor externalities and agency costs. For instance, Kacperczyk et al. (2006) find that unobservable actions consistently predict fund performance. Aside from the liquidity provision style, future research could uncover additional trading characteristics with the estimated daily holdings, such as portfolio pumping, window dressing, and informed trading over short-horizons.

Table 1.1: Intraquarter Holdings MSE

Panel A presents the summary statistics for the mean squared error (MSE) of holdings estimated by several assumptions of trading. GA uses the genetic algorithm to optimize the holdings model developed in this paper. Start, Middle, and End assume that all trades take place on the first, middle, or last day of the quarter respectively. Smooth assumes that the trades take place evenly. For readability, values are scaled by $10^4$. Panel B presents summary statistics for the ratio of MSE for Start, Middle, and End divided to the MSE from the model (GA). The rightmost column contains the p-values for a t-test of whether the log ratio is equal to zero. This tests the null-hypothesis that the ratio is equal to 1. The portfolio allocation error and quarterly MSE are calculated as follows: $Error_t = \sum_{i=1}^{N} \left( \frac{(S_{i,t,M} - S_{i,t,O}) \times P_{i,t}}{TNA_t} \right)^2$ ; $MSE = \frac{1}{T} \sum_{t=1}^{T} Error_t$.

Panel A

| | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| GA | 8.970 | 35.700 | 0.001 | 296.930 |
| Start | 53.980 | 208.510 | 0.007 | 1739.000 |
| Middle | 28.020 | 107.940 | 0.003 | 859.940 |
| End | 59.250 | 223.240 | 0.006 | 1739.600 |
| Smooth | 17.660 | 70.290 | 0.002 | 584.890 |

*Values scaled ($\times 10^4$)

Panel B

| | Mean | Std. Dev. | Min | Max | P-Value (Ratio $\neq$ 1) |
|---|---|---|---|---|---|
| GA/Start | 0.255 | 0.574 | 0.002 | 28.076 | 0.0000 |
| GA/Middle | 0.486 | 0.348 | 0.003 | 6.797 | 0.0000 |
| GA/End | 0.227 | 0.193 | 0.002 | 3.704 | 0.0000 |
| GA/Smooth | 0.781 | 0.439 | 0.005 | 4.863 | 0.0000 |
| Observations | 2693 | | | | |

Table 1.2: Regressions of Holdings MSE on Model Fitness

In Panel A, I regress the intraquarter holdings MSE from the model, estimated with the genetic algorithm, $MSE_{ga}$, onto the model's fitnes, the number of stocks, the number of stocks traded, and the TNA of the fund. In Panel B, I replace $MSE_{calc}$ with the ratio of $MSE_{ga}$ to the error of the next best alternative assumption, $MSE_{smooth}$. All variables are log transformed. Numbers in parentheses are two-way clustered standard errors. *, **, *** denote significance at the 10, 5, and 1% level respectively.

Panel A

| | $ln(MSE_{calc})$ | | | | |
|---|---|---|---|---|---|
| $ln(fitness)$ | -0.420*** | -0.432*** | -0.431*** | -0.405*** | -0.411*** |
| | (0.041) | (0.044) | (0.040) | (0.044) | (0.044) |
| $ln(stocks)$ | | -0.714*** | | | -0.330*** |
| | | (0.119) | | | (0.103) |
| $ln(trade)$ | | | -0.125 | | -0.550 |
| | | | (0.566) | | (0.544) |
| $ln(tna)$ | | | | -0.755*** | -0.559*** |
| | | | | (0.120) | (0.119) |
| Fund & Quarter FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 2,604 | 2,604 | 2,604 | 2,604 | 2,604 |
| $R^2$ | 0.800 | 0.819 | 0.806 | 0.823 | 0.824 |

Panel B

| | $ln(\frac{MSE_{calc}}{MSE_{smooth}})$ | | | | |
|---|---|---|---|---|---|
| $ln(fitness)$ | -0.069*** | -0.075*** | -0.075*** | -0.085*** | -0.076*** |
| | (0.016) | (0.016) | (0.014) | (0.015) | (0.016) |
| $ln(stocks)$ | | 0.450*** | | | 0.457*** |
| | | (0.055) | | | (0.060) |
| $ln(trade)$ | | | -0.216 | | 0.292 |
| | | | (0.254) | | (0.246) |
| $ln(tna)$ | | | | 0.274*** | 0.001 |
| | | | | (0.045) | (0.042) |
| Fund & Quarter FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 2,604 | 2,604 | 2,604 | 2,604 | 2,604 |
| $R^2$ | 0.521 | 0.572 | 0.532 | 0.549 | 0.572 |

Table 1.3: Liquidity Provision Measures (Ancerno Sample)

For each stock on each day, the previous one-day and five-day market-adjusted returns are calculated. The momentum measure, Mom1&5, is defined as the average of the 1 and 5 days market-adjusted returns (Jame, 2018). At the fund level, the momentum measure is calculated quarterly by taking the dollar-volume weighted average $Mom1\&5$ of stocks purchased minus the dollar-volume weighted average of stocks sold. Panel A presents the correlation matrix between the momentum measure estimated based upon various assumptions of daily trades: the observed daily trades (Mom1&5), the algorithm estimated trades($Mom1\&5_{calc}$), the evenly traded ($Mom1\&5_{smooth}$), and the quarterly momentum measure of Jame (2018) (Mom1&5Q). Panel B presents regression results where $Mom1\&5$ is regressed on $Mom1\&5_{calc}$ and $Mom1\&5_{smooth}$. Numbers in parentheses are two-way clustered standard errors. *, **, *** denote significance at the 10, 5, and 1% level respectively.

Panel A: Correlation Matrix

| | $mom1\&5$ | $mom1\&5_{calc}$ | $mom1\&5_{smooth}$ | $mom1\&5Q$ |
|---|---|---|---|---|
| $mom1\&5$ | 1 | | | |
| $mom1\&5_{calc}$ | 0.756 | 1 | | |
| $mom1\&5_{smooth}$ | 0.586 | 0.568 | 1 | |
| $mom1\&5Q$ | 0.599 | 0.583 | 0.989 | 1 |

Panel B: Regression

| | $mom1\&5$ |
|---|---|
| $mom1\&5_{calc}$ | 0.509*** |
| | (0.020) |
| $mom1\&5_{smooth}$ | 0.190*** |
| | (0.021) |
| Observations | 2,604 |
| R-squared | 0.712 |

## Table 1.4: Liquidity Provision Measures (Mutual Fund S12 Sample)

Table 4 presents summary statistics and the correlation matrix of liquidity provision measures estimated for the 59,832 fund-quarters in the S12 sample of active mutual funds from 2002-2016. For each stock on each day, the previous one-day and five-day market-adjusted returns are calculated. Following Jame (2017), the momentum measure, $Mom1\&5$, is defined as the average of the 1 and 5 days market-adjusted returns. At the fund level, the momentum measure $mom1\&5_{calc}$ is calculated quarterly by taking the dollar-volume weighted average $Mom1\&5$ of stocks purchased minus the dollar-volume weighted average of stocks sold. $mom1\&5_{end}$ assumes all trades are made on the final day of the quarter. $Mom1\&5Q$ is estimated quarterly. $\beta_{RLP}$ is the regression-estimated liquidity provision factor of Rinne and Suominen (2014); $Rank_{\beta_{RLP}}$ is the quintile rank of $\beta_{RLP}$.

Panel A: Summary Statistics

| | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| $mom1\&5_{calc}$ | 64,480 | 0.020 | 0.081 | -1.256 | 1.476 |
| $mom1\&5Q$ | 64,480 | 0.004 | 0.050 | -0.630 | 0.510 |
| $mom1\&5_{end}$ | 64,480 | 0.004 | 0.069 | -1.414 | 2.695 |
| $\beta_{RLP}$ | 64,480 | -0.151 | 1.077 | -10.050 | 13.565 |
| $Rank_{\beta_{RLP}}$ | 64,480 | 3 | 13.816 | 1 | 5 |

*Values scaled ($\times$10)

Panel B: Correlation Matrix

| | $mom1\&5_{calc}$ | $mom1\&5Q$ | $mom1\&5_{end}$ | $\beta_{RLP}$ | $Rank_{\beta_{RLP}}$ |
|---|---|---|---|---|---|
| $mom1\&5_{calc}$ | 1 | | | | |
| $mom1\&5Q$ | 0.504 | 1 | | | |
| $mom1\&5_{end}$ | 0.144 | 0.174 | 1 | | |
| $\beta_{RLP}$ | -0.061 | -0.071 | -0.006 | 1 | |
| $Rank_{\beta_{RLP}}$ | -0.056 | -0.071 | -0.004 | 0.760 | 1 |

Table 1.5: 4-Factor Time Series Regressions

Each quarter, I sort the funds into quintiles based on their level of $mom1\&5_{calc}$. I create a new portfolio that is long the portfolio of quintile 5 and short that of quintile 1, rebalancing every month for the length of the time series. I regress the portfolio returns onto the Fama-French 4-factor model. To formulate the portfolios, I used the average measures of $mom1\&5_{calc}$ over the prior two quarters. Column (1) represents the returns to the portfolio long quintile 5 and short quintile 1 of $mom1\&5_{calc}$. Columns (2) and (3) represent returns to the portfolios long quintile 5 and short quintile 1 separately. Numbers in parentheses are standard errors. *, **, *** denote significance at the 10, 5, and 1% level respectively.

| | (1) | (2) | (3) |
|---|---|---|---|
| $\alpha$ | -16.038*** | -0.061 | -16.644*** |
| | (5.078) | (3.883) | (5.434) |
| $R_m - R_f$ | 0.039*** | 0.998*** | 1.037*** |
| | (0.014) | (0.014) | (0.017) |
| SMB | 0.000 | 0.286*** | 0.286*** |
| | (0.030) | (0.023) | (0.028) |
| HML | -0.138*** | 0.014 | -0.124*** |
| | (0.029) | (0.022) | (0.033) |
| UMD | 0.133*** | -0.061*** | 0.072*** |
| | (0.023) | (0.017) | (0.016) |
| $R_{LP}$ | -0.408** | -0.132 | -0.540*** |
| | (0.200) | (0.173) | (0.185) |
| Observations | 165 | 165 | 165 |
| $R^2$ | 0.581 | 0.989 | 0.979 |

Table 1.6: Fama-MacBeth Regressions

I estimate 165 monthly Fama-Macbeth (1973) regressions. Liquidity provision measures, $Mom1\&5_{calc}$, $Mom1\&5_{end}$, $Mom1\&5Q$, and $\beta_{RLP}$ are standardized to have zero mean and unit standard deviation. $Mom1\&5_{calc}$, $Mom1\&5_{end}$, and $Mom1\&5Q$ are averaged over the past two quarters; $\beta_{RLP}$ is estimated over the past two years using monthly returns data (Rinne and Suominen, 2014). Control variables include expenses, turnover, cash, fund age, total net assets, and percent flows. All controls are expressed as percentages, except for TNA and fund age, which are by transformed the natural logarithm. All controls are lagged by one quarter, except fund flows, which are lagged by one, two, and three months. Numbers in parentheses are Newey-West (1987) corrected standard errors. *, **, *** denote significance at the 10, 5, and 1% level respectively.

| | Excess Returns | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $mom1\&5_{calc}$ | -5.608** | | | -4.357** | -5.195*** | -5.917*** |
| | (2.833) | | | (1.719) | (1.608) | (1.634) |
| $mom1\&5Q$ | | -4.840 | | -0.0403 | 2.177 | 2.649 |
| | | (3.091) | | (3.043) | (2.701) | (2.697) |
| $mom1\&5_{end}$ | | | -4.377 | -3.205 | -3.202 | -2.980 |
| | | | (2.836) | (2.601) | (2.390) | (2.444) |
| $Rank_{\beta_{RLP}}$ | | | | | 0.631 | 0.590 |
| | | | | | (1.941) | (1.949) |
| Expenses | | | | | -3.812 | -3.321 |
| | | | | | (5.231) | (5.286) |
| Turnover | | | | | -0.0273 | -0.0230 |
| | | | | | (0.0210) | (0.0214) |
| Cash | | | | | -0.122 | -0.138 |
| | | | | | (0.377) | (0.383) |
| $ln(TNA)$ | | | | | 5.228 | 4.831 |
| | | | | | (3.212) | (3.165) |
| $ln(TNA)^2$ | | | | | -0.510* | -0.454 |
| | | | | | (0.280) | (0.277) |
| $ln(Age)$ | | | | | 29.49** | 33.18** |
| | | | | | (13.41) | (13.85) |
| $ln(Age)$ | | | | | -5.304** | -5.981** |
| | | | | | (2.599) | (2.661) |
| $flow_{t-1}$ | | | | | 0.461 | 0.438 |
| | | | | | (0.380) | (0.381) |
| $flow_{t-2}$ | | | | | -0.184 | -0.135 |
| | | | | | (0.317) | (0.345) |
| $flow_{t-3}$ | | | | | -0.322 | -0.342 |
| | | | | | (0.342) | (0.350) |
| Style FE | No | No | No | No | No | Yes |
| Observations | 120,051 | 120,051 | 120,051 | 120,051 | 120,051 | 120,051 |
| Average $R^2$ | 0.019 | 0.022 | 0.011 | 0.034 | 0.119 | 0.140 |
| Number of months | 165 | 165 | 165 | 165 | 165 | 165 |

Table 1.7: Liquidity Provision and Fund Flows

I estimate pooled cross-sectional and time series regressions of monthly fund flows onto the $mom1\&5_{calc}$ with controls for performance. Controls follow Del Guercio and Tkac (2002) and include: an indicator for outperforming the S&P 500, lagged excess returns (separated above and below the S&P 500), Jensen's alpha (separated above and below the S&P 500), and the tracking error (separated above and below the Jensen's alpha above the lagged excess returns above the S&P 500), fund age, TNA, lagged flow, and year × style dummies. Due to the coefficient magnitude, I scale $mom1\&5_{calc}$ by 1000. In the first regression column, flows are measured in millions of dollars; the second column uses a percent flow expressed as fraction between 0 and 1. Standard errors are clustered by fund and month. *, **, *** denote significance at the 10, 5, and 1% level respectively.

| | Flow ($) | Flow (%) |
|---|---|---|
| $mom1\&5_{calc}$ (× 1,000) | 1,267 | -0.028 |
| | (1,010) | (0.235) |
| Outperform S&P 500 dummy | 2.382** | 0.001 |
| | (0.983) | (0.001) |
| Lagged excess return (above S&P 500) | 49.657 | 0.226*** |
| | (36.792) | (0.032) |
| Lagged excess return (below S&P 500) | 125.699*** | 0.130*** |
| | (35.596) | (0.0321) |
| Jensen's alpha (above S&P 500) | 1,900*** | 1.587*** |
| | (264.826) | (0.113) |
| Jensen's alpha (below S&P 500) | 1,090*** | 0.900*** |
| | (181.182) | (0.122) |
| Tracking error (above S&P 500) | 10.55*** | 0.005** |
| | (3.645) | (0.002) |
| Tracking error (below S&P 500) | -0.624 | -0.021*** |
| | (2.901) | (0.003) |
| Controls | Yes | Yes |
| Observations | 120,051 | 120,051 |
| $R^2$ | 0.143 | 0.028 |

Figure 1.1: Distribution of Fitness: Ancerno Sample

Figure 1 plots the frequency distribution of fitness values for the sample of 2,604 fund quarters in Ancerno. I use the genetic algorithm to minimize equation (13) for each fund-quarter in order to find the dynamic portfolio that best replicates fund's observed daily returns:

$$\min_{k_1 \ldots k_N} \sum_{t=1}^{T} \left[ \left( R_{t,F} - \sum_{i=1}^{N} R_{i,t} \times \frac{S_{i,k_i,t-1} \times P_{i,t-1}}{\sum_{i=1}^{N} S_{i,k_i,t-1} \times P_{i,t-1}} \right)^2 + 0.01 V_t \right].$$

$R_{t,F}$ is the return to the fund on day $t$. $R_{i,t}$ and $P_{i,t}$ are the return and price of stock $i$ on day $t$. $S_{i,k,t-1}$ is the number of shares held in stock $i$ on day $t-1$. $S_{i,k,t-1}$ takes on a value equal to $S_T$ (the end of quarter holdings) if $k \geq t-1$ and $S_0$ (the start of quarter holdings) otherwise. There are $N$ stocks in the portfolio and T days in the quarter. $V_t$ is an indicator variable equal to 1 if the implied flows constraint is violated and zero otherwise. I minimize the function by selecting values of $(k_1, \ldots, k_N)$. For ease of interpretation, I refer to the fitness of the function as the evaluation of equation (13) multiplied by -1. This allows for higher values of fitness to be interpreted as a greater fit on the data.

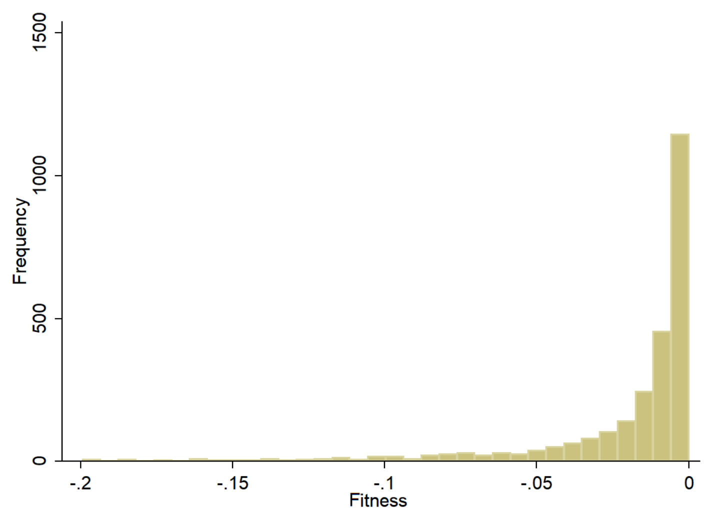


Figure 1.2: $Mom1\&5_{calc}$ Persistence

This figure presents the average values of $mom1\&5_{calc}$ based on quintile sorts of the fund's average $mom1\&5_{calc}$ over the past two quarters from 2002Q4 to 2016Q2. Following Jame (2017), the momentum measure, $Mom1\&5$, is defined as the average of the 1 and 5 days market-adjusted returns. At the fund level, the momentum measure $mom1\&5_{calc}$ is calculated quarterly by taking the dollar-volume weighted average $Mom1\&5$ of stocks purchased minus the dollar-volume weighted average of stocks sold.

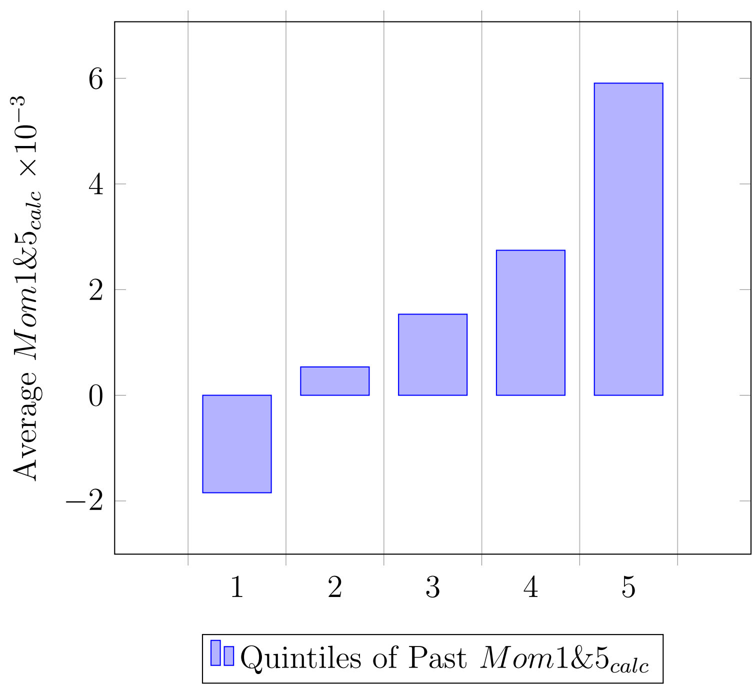
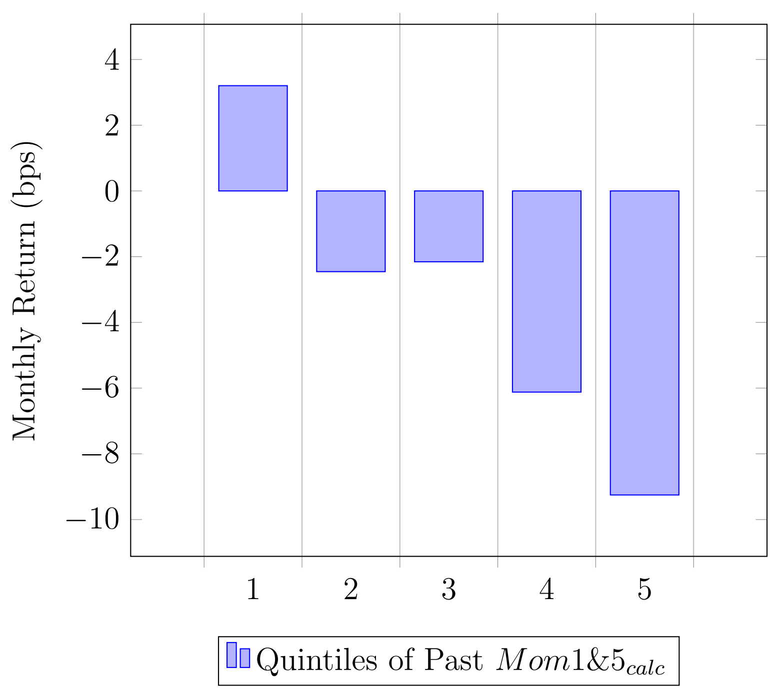


Figure 1.3: $Mom1\&5_{calc}$ and Future Returns

Each quarter, I sort the funds into quintiles based on their level of $mom1\&5_{calc}$ over the past two quarters. Following Jame (2017), the momentum measure, $Mom1\&5$, is defined as the average of the 1 and 5 days market-adjusted returns. At the fund level, the momentum measure $mom1\&5_{calc}$ is calculated quarterly by taking the dollar-volume weighted average $Mom1\&5$ of stocks purchased minus the dollar-volume weighted average of stocks sold. This figure presents the average market excess monthly returns returns for the five portfolios from October 2002 to June 2016.



# Chapter 2 The Democratization of Investment Research: Implications for Retail Investor Profitability and Firm Liquidity

## 2.1 Introduction

Information is a key ingredient for well-functioning financial markets. Without a broad base of investors with access to accurate information, pricing securities becomes difficult and markets can stagnate. At the same time, information with high investment value tends to be costly to produce, which has left individuals at a perennial disadvantage relative to institutional investors. In recent decades, improvements in technology have significantly reduced the cost of gathering and sharing investment information, and these developments have been lauded for their potential to help level the informational playing field.[1]

In this article, we examine the extent to which an important technology-enabled innovation, crowdsourced investment research, has led to improved individual investor decision-making and enhanced market liquidity. Few finance media sites exemplify the democratization of investment research better than Seeking Alpha (SA). Seeking Alpha attracts millions of visitors each month by providing curated investment research from thousands of individual research contributors.[2] The value of Seeking Alpha's investor-authored research is well documented. For example, Chen, et al. (2014) find that Seeking Alpha research articles predict future stock returns and earnings surprises. In our analysis, we examine whether individual investors benefit from Seeking Alpha research.

Individual investors have been traditionally viewed as unsophisticated "noise" traders who tend to spend less time on investment analysis, use different information sources from their professional counterparts, and underperform standard benchmarks (e.g., Kumar and Lee, 2006; Barber and Odean, 2013 review the literature). In recent years, however, studies have uncovered evidence of informed trading by individuals, with retail order flow predicting stock returns and future earnings announcement surprises (Kaniel, Saar, and Titman, 2008; Kaniel et al., 2012; Kelley and Tetlock, 2013, 2017; Boehmer, Jones, and Zhang, 2017). Although the improved performance of individual investors over time could be associated with learning or changing demographics, a key driver that remains unexplored is better access to investment research.

We begin by documenting that Seeking Alpha's crowdsourced investor-authored research is distinct from traditional Wall Street brokerage research. We analyze over 140,000 research articles for 4,000 stocks and find that after controlling for other firm characteristics, Seeking Alpha coverage (number of articles) is higher among

[1]For example, early in the internet era SEC Commissioner Laura Unger anticipated technology's potential and concluded a speech with "It looks as though investors stand to benefit greatly from the Information Revolution. The Internet has powered the revolution. It's also been a key element in the push for democratization of the flow of investment information." (June 2000) https://www.sec.gov/news/speech/spch387.htm

[2]https://seekingalpha.com/page/about_us

firms with low institutional ownership and greater breadth of ownership, whereas the opposite is true for brokerage research coverage. The differences highlight Seeking Alpha's emphasis on retail investor oriented research.

We find strong evidence that individual investors react to Seeking Alpha research. Using trade and quote data from NYSE TAQ and the method of Boehmer, Jones, and Zhang (2017) to identify retail investor trades, we find a significant increase in retail trading on days with Seeking Alpha articles. Individuals also account for a greater fraction of overall trading volume on article days, suggesting that individuals react to Seeking Alpha research more than institutional investors do. In addition, we find that retail order imbalances are strongly correlated with the sentiment of Seeking Alpha research articles, with individual investors in aggregate being more likely to purchase (sell) stock following positive (negative) articles.

More importantly, we find robust evidence that access to crowdsourced research enhances the profitability of retail investor trades. In particular, the relation between retail order flow and future stock returns is roughly twice as strong on days with Seeking Alpha research articles. For example, a one standard deviation increase in daily retail order imbalance is associated with future ten-day returns that are 0.09% larger on average, and yet this return differential increases to 0.18% on days with Seeking Alpha articles and to 0.34% on days with articles written by high-skill research contributors. We find no evidence of a return reversal, which is consistent with individual investors becoming informed through their access to crowdsourced research.

We benchmark retail investor behavior on days with Seeking Alpha articles against their behavior on days with brokerage research. Although retail investors may have access to consensus recommendations or forecasts, their access to detailed reports and to the analysts themselves is less extensive than institutional investors. Accordingly, we find weaker evidence of increased retail trading around days with brokerage research revisions. Moreover, we find no evidence that retail investors trade profitably around brokerage research revision days, indicating the limited value of brokerage research to retail investors and highlighting the distinctiveness of crowdsourced investment research. We also find no evidence that retail investors trade more profitably following stock-focused media articles, further underscoring the investment value of Seeking Alpha.

Notably, we find the opposite pattern when examining how institutional investors trade following these events. In particular, we find that institutional order flow is more informed following brokerage research and media articles, consistent with superior access to sell-side research and a greater capacity for processing public information. On the other hand, we do not observe that institutional order flow is more informative following Seeking Alpha articles, suggesting that crowdsourced investment research provides unique information benefits to retail investors.

We next we explore the broader implications of crowdsourced research for the firm's information environment. We hypothesize that Seeking Alpha research reduces information asymmetry among investors by decreasing the information advantage of institutional investors over retail investors, and we test whether the reduction in information asymmetry is significant enough to translate into improved liquidity for the firm. Our empirical strategy is to identify plausibly exogenous shocks to Seeking Alpha coverage (similar to Kelly and Ljungqvist, 2012). Specifically, we contend that when an individual contributor departs from the platform altogether, the resulting decline in Seeking Alpha coverage is exogenous, i.e. driven by a change in the contributor's personal circumstances rather than an expectation of a change in the firm's information environment. Consistent with the view that Seeking Alpha contributor departures are exogenous events, we find no evidence that firms that experience departures also experience reductions in coverage by non-departing Seeking Alpha contributors, brokerage firms, or the media.

We estimate the effects of an exogenous reduction in Seeking Alpha coverage on firm-level measures of liquidity in the year following the event using a difference-in-difference approach. Specifically, we define a firm as having experienced contributor departure when 20% or more of the firm's contributors depart Seeking Alpha in a given year, and we match each treated firm to control firms that are in the same size and book-to-market quintiles and experience no contributor departure. We find that bid-ask spreads and the Amihud (2002) illiquidity measure increase by 2.4% and 5.5%, respectively, for contributor-departure firms relative to control firms, which is consistent with the idea that reduced Seeking Alpha coverage results in lower liquidity. The effects of contributor departures on liquidity are stronger among small firms, those with greater retail ownership, and when the departing contributors possess greater skill.

Our study contributes to a nascent but fast-growing literature on the role of crowdsourced research in capital markets. We extend early studies that document the role of crowdsourced research in predicting future returns and earnings (Chen et al., 2014; Jame et al., 2016; Avery, Chevalier, and Zeckhauser, 2016) by documenting its role as a source of information specifically for retail investors. Our findings that Seeking Alpha research encourages retail investor participation and helps retail investors become more informed are consistent with crowdsourced research levelling the informational playing field between institutional and retail investors.

Our analysis also adds to the literature that studies the performance of retail investors. Early studies find the trading performance of retail investors to be subpar due to behavioral biases, lack of sophistication, and poor access to information with high investment value (e.g., Barber and Odean, 2000, Kumar and Lee, 2006; Frazzini and Lamont, 2008; Hvidkjaer, 2008; Barber, Odean, and Zhu, 2009). On the other hand, more recent work finds that retail investors as a group exhibit stock picking ability and speculate that retail investors have valuable information gleaned from geographic proximity to firms, relations with employees, or insights into customer tastes (e.g. Kaniel, Liu, Saar, and Titman, 2012; Kelley and Tetlock, 2013, 2017; and Boehmer, Jones, and Zhang, 2017). Our study is distinguished by its focus on how retail investors become informed. We present results that link retail investors' trading performance to the availability of retail-oriented investment research and question the view that retail investors make investment decisions without requisite investment analysis.

A third stream of literature examines the use of technology by regulators to level the informational playing field between institutional investors and retail investors.[3] We complement these studies by examining the extent to which a technology-enabled market innovation, Seeking Alpha, has democratized the flow of investment information. Our findings illustrate how technological change can enable new business models which improve retail investors' access to information and level the informational playing field among investors.

### 2.2 The Seeking Alpha Sample

Seeking Alpha (SA) is one of the largest investment-related social media websites in the US. In 2017, the site had more than 39 million monthly visits, with the average visit lasting roughly 20 minutes (Seeking Alpha, 2018). The website relies on a contributor network of over 15,000 individuals to publish opinion articles. Contributor testimonials suggests that some of the primary motivations for contributing articles include direct compensation from Seeking Alpha, feedback (via reader comments) on investment theses, and increased recognition and visibility which may lead to other professional opportunities.[4] Chen, et al. (2014) find that Seeking Alpha's crowdsourced investment research contains valuable information, with articles and user commentaries predicting future stock returns and earnings surprises.

We collect all opinion articles published between 2005 and 2017 on the Seeking Alpha website. For each article, we collect the following information: article ID (assigned by Seeking Alpha), title, main text, date of publication, author name, and ticker (or tickers) assigned to each article. Following Chen et al., (2014) we limit the sample to articles that are associated with one ticker. We further limit the sample to common stocks (share codes 10 and 11) with available data in the CRSP-Compustat merged database. Our final sample includes 156,513 single-ticker articles.

For each firm, we collect data on share price, shares outstanding, stock returns, volume and closing bid and ask prices from CRSP. We obtain book value of equity, book value of debt, book value of assets, the date of the initial public offering (IPO), earnings before interest taxes depreciation and amortization (EBITDA), and total common shareholders from Compustat. We collect the number of shares held by institutions from the Thomson Reuters Institutional Holdings (S34) database. We obtain earnings announcement dates and sell-side analyst earnings forecast from the IBES unadjusted US detail history file, sell-side analyst recommendations from the IBES detail recommendation file, and earnings guidance from the IBES detail history guidance file. Data on traditional media coverage, defined as Dow Jones News Service articles, were graciously provided by Byoung-Hyoun Hwang for the period from 2005

[3] Examples are the launch of the Electronic Data Gathering, Analysis, and Retrieval system (EDGAR) in 1993 (Asthana, Balsam, and Sankaraguruswamy, 2004; Gao and Huang, 2017), Regulation Fair Disclosure in 2000 (Eleswarapu, Thompson, and Venkataraman, 2004; Duarte, Han, Harford, and Young, 2008), and the mandated use of eXtensible Business Reporting Language (XBRL) in corporate filing in 2009 (Blankespoor, Miller, and White, 2014; Bhattacharya, Cho, and Kim, 2018).

[4] See: https://seekingalpha.com/page/testimonials.

to 2012 (as in Chen et al, 2014), and we collect the data for the period from 2012 to 2017.

Table 2.1 describes the remarkable increase in the breadth and depth of Seeking Alpha coverage over time. From 2005 to 2017, research coverage rose from 240 to 2,217; the number of research contributors grew from 38 to 1,995; and the number of research articles increased from 282 to 19,505. In an average year, 1,208 unique contributors publish 12,039 articles on 1,638 different companies. Conditional on having Seeking Alpha coverage, the average firm has roughly 6.6 articles per year, written by 4 different contributors.

## 2.3 Contrasting Seeking Alpha with Traditional Brokerage Research

In this section, we explore the attributes of Seeking Alpha research in relation to traditional Wall Street research. Seeking Alpha's business model is built on reaching a wide audience of do-it-yourself investors, and Seeking Alpha contributors are often individual investors. In contrast, prior survey evidence and empirical work suggests that brokerage analysts cater to institutional investors. For example, Brown et al. (2015) report that more than 80% of surveyed analysts view hedge funds and mutual fund clients as very important, while only 13% of these analysts view retail clients as important. Consistent with the survey evidence, several papers find that sell-side research is strongly increasing in institutional ownership (see, e.g., Bhushan, 1989; Green et al., 2014).

We examine the determinants of Seeking Alpha coverage and sell-side coverage by estimating the following panel regression:

$$\begin{aligned} Coverage_{it} = {} & \alpha + \beta_1 InstitutionalOwnrshp_{i,t-1} + \beta_2 BreadthofOwnrshp_{i,t-1} \\ & + \beta \times Chars + Time_t + \epsilon_{it} \end{aligned} \tag{2.1}$$

where *Coverage* is the natural log of 1 plus the total number of unique Seeking Alpha contributors writing at least one article for the stock during the calendar year (*SA Coverage*), or the natural log of 1 plus the total number of unique brokerage firms issuing at least one earnings forecast for the stock during the calendar year (*IBES Coverage*).

The two independent variables of primary interest are *Institutional* $Ownrshp_{i,t\text{-}1}$, defined as the percentage of the firm's shares held by institutional investors in year *t*-1, and *Breadth of Ownrshp*, defined as the number of common shareholders (both in logs). The vector of firm characteristics, *Chars,* includes: market capitalization (*Size*), book to market (*BM*), return volatility (*Volatility*), share turnover *(Turnover)*, past one-year return (*Return*), past one-year profitability (*Profitability*), and the number of unique media articles mentioning the firm the prior year (*Media Coverage*). See the Appendix for detailed definitions. We log all continuous variables other than *Profitability* and *Return*, and standardize all variables to have zero mean and unit variance. We include year fixed effects and cluster standard errors by firm.

Specification 1 of Table 2 examines the determinants of *SA Coverage* without controlling for *IBES Coverage.* In general, *SA Coverage* is higher for larger firms,

firms with more frequent media coverage, and those with greater trading volume. In addition, *SA Coverage* is positively related to volatility, past one-year returns, and profitability. Consistent with our conjecture that Seeking Alpha research is a retail investor rather than an institutional investor phenomenon, we find a strong negative relation between *SA Coverage* and institutional ownership and a strong positive relation between *SA Coverage* and total common shareholders. In particular, a one standard deviation increase in *Institutional Ownership* (*Breadth of Ownership*) is associated with an 18% decline (3% increase) in *SA Coverage*. These findings are robust to controlling for *IBES Coverage*, which is positively correlated with *SA Coverage* (Specification 2).

Specifications 3 and 4 present analogous results for traditional coverage by sell-side brokerage firms (*IBES Coverage*). As expected and in sharp contrast to the *SA Coverage* patterns, *IBES Coverage* is strongly positively related to institutional ownership and strongly negatively related to breadth of ownership. Collectively, these results are consistent with the idea that Seeking Alpha tilts their research towards stocks with greater retail ownership and a larger investor base, while traditional brokerage research caters to institutional investors.

### 2.4 Do Retail Investors React to Seeking Alpha Research?

In this section, we examine two related predictions: 1) the dissemination of Seeking Alpha research generates disproportionately greater trading among retail investors (Section 2.4.1), and 2) the tone of the Seeking Alpha research article influences retail investor order imbalances (Section 2.4.2).

#### 2.4.1 Retail Trading Intensity around Seeking Alpha Research

We identify retail trading using Boehmer, Jones, and Zhang's (2017) approach, which exploits two key institutional features of retail trading. First, most equity trades by retail investors take place off-exchange, either filled from the broker's own inventory or sold by the broker to wholesalers (Battalio, Cowin, and Jennings, 2016). TAQ classifies these types of trades with exchange code "D." Accordingly, we identify retail trades by limiting our analysis to trades executed on exchange code "D."

Second, retail traders typical receive a small fraction of a cent price improvement over the National Best Bid or Offer (NBBO) for market orders (ranging from 0.01 to 0.2 cents), while institutional orders tend to be executed at whole or half-cent increments. Thus, we follow Boehmer, Jones, and Zhang (2017) (BJZ) and identify trades as retail purchases (sales) if the trade took place at a price just below (above) a round penny.[5] The BJZ approach is conservative in the sense that it has a low type 1 error (i.e., trades classified as retail are very likely to be retail). While this approach does omit some retail trading including nonmarketable limit orders and

[5]This approach focuses on liquidity-demanding retail trading (market orders). Kelley and Tetlock (2013) find that while both aggressive and passive retail trading predicts returns, liquidity-demanding trading also predicts earnings surprises.

retail traders that take place on registered exchanges, it "picks up a majority of overall retail trading activity" (BJZ page 6).[6]

We consider two complementary measures of retail trading intensity: *Retail Turnover*$_{it}$, which is the total trading volume in stock $i$ on day $t$ classified as retail, scaled by stock $i$'s shares outstanding, and *Percentage Retail Turnover*, which is the total trading volume in stock $i$ on day $t$ classified as retail, scaled by aggregate trading volume for stock $i$ on day $t$. The turnover measure allows us to examine whether retail investors trade more following Seeking Alpha research than at other times, whereas the percentage turnover measure considers whether the increase in trading following SA research is larger for retail investors than for institutional investors. We estimate the effects of Seeking Alpha research on retail investor trading using the following daily panel regression:

$$\begin{aligned} RetailTrading_{it} = \alpha + \beta_1 SA_{it-1,t} + \beta_2 IBES_{it-1,t} + \beta_3 Media_{t-1,t} \\ + \beta_4 Chars_{iy-1} + Time_t + Firm_i + \epsilon_{it} \end{aligned} \tag{2.2}$$

where *Retail Trading* is *Retail Turnover* or *Percentage Retail Turnover*, $SA_{it}$ is equal to one if Seeking Alpha issued a research report for firm $i$ on day $t$ or *t-1*,[7] and *IBES* and *Media* are defined similarly. *Char* includes *Size*, *book-to-market* (*BM*), *Institutional Ownership*, *Volatility*, *Turnover*, *Return*, *Profitability*, *IBES coverage*, and *Media Coverage*, all measured at the end of the previous year. Detailed variable descriptions are provided in the Appendix. All continuous independent variables are standardized to have mean zero and unit variance. *Time* and *Firm* indicate time (calendar day) and firm fixed effects. Standard errors are clustered by firm.

Our TAQ sample begins in 2007, and we consider all firm-years with Seeking Alpha coverage (at least one SA research report). We exclude dates that coincide with earnings announcements or earnings guidance to reduce the likelihood that SA research is merely proxying for information contained in these information events. The resulting 2007-2017 sample comprises 3,925,070 firm-day observations.

Specifications 1 and 3 of Table 2.3 report the results prior to including firm fixed effects. The statistically significant positive coefficient on $SA_{t\text{-}1,t}$ in Specification 1 indicates that retail investors trade more following Seeking Alpha research than on other days. Moreover, the significant positive coefficient in Specification 3 confirms that the increase in trading for retail investors is greater than for institutional investors, with retail investors comprising a larger fraction of the daily turnover following Seeking Alpha research. The results also hold after including firm fixed effects in Specifications 2 and 4.

Table 2.3 also provides evidence that trading intensity is greater following the release of traditional brokerage research or media coverage, although the economic

[6]BJZ also note that in a conference discussion of their work, Eric Kelley presented that the correlation between the BJZ order imbalance measure and imbalances calculated from Kelley and Tetlock (2013)'s proprietary retail data with observed trade directions is in the range of 0.345 to 0.507, with an average of 0.452.

[7]We include day $t$-1 to account for SA reports published after the close of trading on day $t$-1. We analyze trading days separately in Table 2.4.

magnitudes are considerably smaller. For example, the coefficient on the *SA* indicator variable in Specification 3 (1.31%) is roughly 4 times larger than the coefficient on the *IBES* indicator (0.34%) and more than 50% larger than *Media* (0.69%), and the relative differences become even larger after controlling for firm fixed effects.

#### 2.4.2 Retail Order Imbalances and Seeking Alpha Research Article Sentiment

To provide more direct evidence that retail investors read and react to Seeking Alpha research, we explore the relation between retail order imbalances and SA research sentiment. We consider two sentiment measures. The first is based on article tone: *Percent Negative*$_{it}$ (*Percent Positive*$_{it}$) is the average fraction of negative (positive) words across all single-ticker articles published on Seeking Alpha about company $i$ on day $t$ (Chen et al., 2014). We use the word list compiled by Loughran and McDonald (2011) to classify words as negative or positive. The second is based on a contributor's investment position, as disclosed in the article: *Short (Long) Position*$_{it}$ is one when a contributor discloses short (long) the stock (Campbell et al., 2016).[8] When several contributors disclose investment positions, we take a simple average.

For reference, we also examine the relation between retail order imbalances and brokerage research and traditional media coverage. We create indicator variables for *Long* (*Short*) brokerage research, defined as IBES recommendation upgrades (downgrades) or positive (negative) earnings forecast revisions. We also calculate the tone of traditional media articles using the *Percent Negative*$_{it}$ and *Percent Positive*$_{it}$ measures, as defined above.

For all firm-days with at least one event (i.e., Seeking Alpha article, IBES report, or media article) we estimate the following panel regression:

$$\begin{aligned} RetailOIB_{i,t+x} = {} & \alpha + \beta_1 SA + \beta_2 SA \times PercentNegative_{it} \\ & + \beta_3 SA \times PercentPositive_{it} \beta_4 SA \times ShortPosition_{it} \\ & + \beta_5 SA \times LongPosition_{it} + \beta_6 IBES \\ & + \beta_7 IBES \times Long + \beta_8 IBES \times Short + \beta_9 Media \\ & + \beta_{10} Media \times PercentNegative_{it} \\ & + \beta_{11} Media \times PercentPositive_{it} \\ & + \beta_{12} InstOIB_{it} + \beta_{13} Char_{iy-1} + Time_t + \epsilon_{it}. \end{aligned} \tag{2.3}$$

As in Boehmer, Jones, and Zhang (2017), *Retail OIB*$_{i,t+x}$ is defined as the retail buy volume less the retail sell volume, scaled by the total retail trading volume for firm $i$ on day $t+x$. We define day $t$ as the event day (the day the Seeking Alpha article is published) and let $x$ vary from -3 to +3. By focusing on a seven-day window around the event, we are able to examine the lead-lag relation between Seeking Alpha research and retail investor trading. To control for broad omitted factors that affect the order imbalances of all investors, we also include *Inst OIB*,

[8]The site has mandated investment position disclosures since 2015.

which is defined as the total order imbalance across all TAQ trades less the retail order imbalance.[9] *Char* is a vector of firm characteristics (taken from BJZ) and includes past one-week returns ($Ret_{w\text{-}1}$), past one month returns ($Ret_{m\text{-}1}$), returns over the prior two to seven months ($Ret_{m\text{-}7,m\text{-}2}$), market capitalization (*Size*), monthly turnover (*Turnover*), monthly volatility of daily returns (*Volatility*), and book-to-market (*BM*). With the exception of returns, all control variables are measured at the end of the previous year and are in natural logs. To control for confounding events, we exclude Seeking Alpha articles that coincide with earnings announcements, earnings guidance, or brokerage (IBES) investment research (forecast revisions or recommendation changes).

Table 2.4 presents the results. We find robust evidence that Seeking Alpha article sentiment tone predicts retail order imbalances on event days 0 and +1. For example, a one-standard deviation increase in *SA Percent Negative Words* is associated with a 1.17 % and 1.05% decline in *Retail OIB* on days 0 and 1.[10] On these days, the incremental effect of short positions is a decline in *Retail OIB of* 6.75% and 6.41%, respectively, whereas the incremental effect of long positions is 3.83% and 2.15%. Short position disclosures and negative SA tone continue to have a discernible effect on Retail OIB on days +2 and +3. We find no evidence that retail order imbalances on event days -3 or -2 predicts the sentiment of future Seeking Alpha articles, although there is some evidence that *Retail OIB* on day -1 is related to negative article tone and the disclosure of short positions.

We also find that retail investors react to IBES research and media coverage, yet the economic magnitudes of these effects are smaller than for Seeking Alpha research. For example, while the cumulative effects of *SA*×*Long* and SA×Short on *Retail OIB* in the two days surrounding the event are 5.98% and -13.16%, the corresponding estimates for *IBES*×*Long* and *IBES*×*Short* are 4.22% and -1.03%. Moreover, the effect of *SA Percent Negative* on retail order imbalances in the same event window is twice that of *Media Percent Negative* (-2.22 versus -1.19).

In summary, our findings thus far indicate that Seeking Alpha research is geared towards stocks that are owned by retail investors. Moreover, the publication of SA articles stimulates significant trading by retail investors that is directionally consistent with the tone of the article. Given the evidence that Seeking Alpha research tone is informative about future stock returns (Chen et al., 2014), these findings suggest that Seeking Alpha research may also facilitate more informed trading among retail investors. We explore this hypothesis next.

[9] Our trade-based classification approach provides an imperfect measure of all retail trading, and therefore calculating the total order imbalance less the retail order imbalance also provides an imperfect measure of all institutional trading.

[10] The absence of a relation between *Retail OIB* and *SA Percent Positive Words* is consistent with prior evidence that negative words are better at capturing variation in tone (e.g., Tetlock, 2007).

## 2.5 Does Seeking Alpha Research Help Retail Investors Become Better Informed?

We examine the informativeness of retail order imbalances on days in which Seeking Alpha research is published by estimating the following panel regression:

$$
\begin{aligned}
Ret_{it,t+x} = \alpha &+ \beta_1 RetailOIB_{it} + \beta_2 InstOIB_{it} + \beta_3 SA_{it} + \beta_4 RetailOIB \times SA_{it} \\
&+ \beta_5 InstOIB_{it} \times SA_{it} + \beta_6 Media_{it} + \beta_7 RetailOIB \times Media_{it} \\
&+ \beta_8 InstOIB \times Media_{it} + \beta_9 IBES_{it} + \beta_{10} RetailOIB \times IBES_{it} \\
&+ \beta_{11} InstOIB \times IBES_{it} + \beta_{13} Char_{iy-1} + Time_t + \epsilon_{it}
\end{aligned}
\tag{2.4}
$$

where $Ret_{it+1,t+x}$ is the return on stock $i$ from the close of day $t$ to the close of day $t$+$x$, with $x$ equal to 1, 5, or 10; *Retail OIB* is the total retail buy volume less the total retail sell volume, scaled by the total retail trading volume; *Inst OIB* is the total non-retail buy volume less the total non-retail sell volume, scaled by the total non-retail trading volume; and $SA_{it}$ is equal to one if there is a SA research article on firm $i$ on day t or *t-1*, and zero otherwise. Our primary variable of interest, *Retail OIB*×*SA*, captures the incremental informativeness of retail order imbalance on days with Seeking Alpha research. *Institutional OIB*×*SA* captures the incremental informativeness of institutional trading on days with Seeking Alpha research. $Media_{it}$ is equal to one if there is a traditional media article for firm $i$ on day t or *t-1*, and zero otherwise. Similarly constructed, $IBES_{it}$ indicates the distribution of sell-side research. *Characteristics* is the vector of firm characteristic described in Equation (2.3). We exclude firm-days that coincide with earnings announcements or earnings guidance and standardize all continuous variables, as in Section 4. We include time fixed effects and double cluster standard errors by date (calendar day) and firm.

Table 2.5 presents the results. Consistent with Boehmer, Jones, and Zhang (2017), we find that retail order imbalance is a strong positive predictor of future returns. Moreover, we find that retail trading is a stronger return predictor on days with Seeking Alpha research. For example, in Specification 3, a one standard deviation increase in retail order imbalance is associated with a 0.09% increase in 10-days returns on days without Seeking Alpha research and an increase of 0.18% (0.09% +0.09%) on days with Seeking Alpha research. In contrast, the coefficients on *Retail OIB*×*Media* and *Retail OIB*×*IBES* are always economically and statistically insignificant, which further highlights the unique role of Seeking Alpha research (relative to traditional media and brokerage research) in enhancing the informativeness of retail investor trades. We also find that when retail investor trading is more informed, institutional investor trading tends to be less informed. Specifically, *Institutional OIB* is a stronger predictor of future returns on days when traditional media articles and sell-side research are distributed but not on days when SA articles are published.

Finally, in Figure 1 we estimate Equation (2.3) for $x$=10, 20, and 60 and plot the estimates of *Retail, Retail OIB*×*SA, Retail OIB*×*Media,* and *Retail OIB*×*IBES.* Consistent with Boehmer, Jones, and Zhang (2017), we find that the informativeness of retail trading is concentrated over relatively short holding periods. In each case, more than 50% of the 60-day returns accrue in the first ten days. The patterns are similar but economically smaller for non-event days and days with media articles or IBES research. The lack of reversal is inconsistent with the view that returns following retail trading reflect uninformed price pressure.

### 2.5.1 Retail Order Imbalances and Stock Returns: Robustness

In the Internet Appendix (Table IA.1), we confirm that the evidence regarding the incremental informativeness of retail order imbalances around Seeking Alpha research is robust to several alternative methodological choices including: (1) measuring retail order imbalances using number of trades instead of share volume; (2) including in the sample firm-days with earnings news (earnings announcements or earning guidance); (3) adding firm fixed effects to the panel regression; and (4) using Fama-MacBeth regressions to estimate Equation (2.4).[11]

We explore whether the results are stable over time by estimating Equation (2.4) each month (Specification 3 in Table 2.5) and we plot the cumulative coefficients on *Retail OIB*$\times$*SA* in Figure IA.1. We observe a jump in the second half of 2008, consistent with SA research being particularly valuable during the financial crisis, and a fairly stable and positive drift over the full sample period. To confirm that our results are not driven by the financial crisis period, we re-estimate the model after excluding the second half of 2008, and continue to find that the coefficient on Retail OIB $\times$ SA is statistically significant at a 5% level.

### 2.5.2 Conditioning on Firm Size and Contributor Skill

We examine whether Seeking Alpha research is more useful to retail investors for smaller stocks, which are known to be less informationally efficient. Specifically, we sort firms into two groups based on the median breakpoint of market capitalization (measured at the end of the previous year). We then repeat Specification 3 of Table 2.5 for each size group separately. Panel A of Table 2.6 reports the coefficient on *Retail OIB*$\times$*SA* for each size subgroup. Among smaller stocks, the incremental profitability of retail order flow associated with Seeking Alpha research is 0.17 % over a 10-day holding period and statistically significant, compared to the 0.09% coefficient on *Retail OIB (*unreported*)*. In contrast, among larger stocks the incremental informativeness of retail trading on days with Seeking Alpha research is considerably smaller. For example, the coefficient on *Retail OIB* over a 10-day horizon is 0.02% and the coefficient on *Retail OIB*$\times$*SA* is 0.04%, neither of which is statistically different from zero. The findings suggest that the impact of Seeking Alpha research on the informativeness of retail trading is concentrated among smaller stocks.

We suggest that retail investors will benefit more from research written by skilled Seeking Alpha contributors. Motivated by the idea that research by skilled individuals

[11] In estimating Fama-Macbeth regressions, we limit the sample to days with at least 10 firms with Seeking Alpha research, and we calculate Newey West standard errors with 1, 5, and 10-day lags depending on the return horizon.

has greater market impact, we proxy for contributor skill by averaging the two-day (0, 1) absolute market-adjusted returns across the last five articles written by the contributor. We partition the sample into high versus low skill based on the median breakpoint of contributor skill, and we use indicator variables to separate event days where the fraction of articles written by high skill contributors is greater than 50% (*High Skill*) from those where it is less than or equal to 50% (*Low Skill*).

The results are presented in Panel B of Table 2.6. The coefficients on *Retail OIB*$\times$*SA High Skill* are highly significant, ranging from 0.11% for a one-day event window to 0.25% for the 10-day window. In contrast, the coefficients on *Retail OIB*$\times$*SA Low Skill* are small and statistically insignificant. Furthermore, across all holding periods, the difference between the two estimates is significantly different from zero. The link between contributor skill and the extent to which retail order imbalance is informative about future returns is consistent with more informed investment research leading to more informed retail trading.

## 2.6 Seeking Alpha Research, Information Asymmetry, and Firm Liquidity

The evidence in the previous section suggests that Seeking Alpha disseminates investment information to retail investors who would not otherwise have access to the information. If crowdsourced investment research leads to a material reduction in the level of information asymmetry between institutions and individuals, then Seeking Alpha coverage should improve firm liquidity. In this section, we explore the broader implications of crowdsourced research for the firm's information environment. Our approach focuses on exogenous shocks to Seeking Alpha coverage, and we study the effects of declines in coverage on measures of firm liquidity.

### 2.6.1 Identifying Exogenous Shocks to Seeking Alpha Coverage

Studying the effects of Seeking Alpha coverage on firm liquidity is difficult because SA contributors choose which stocks to cover, and the choice to cover a stock is likely influenced by many firm characteristics including liquidity itself. We attempt to circumvent this challenge by identifying changes in Seeking Alpha coverage that are unlikely to be driven by firm characteristics. Specifically, our identification strategy focuses on the departure of a Seeking Alpha contributor from Seeking Alpha. Our underlying assumption is that when a contributor leaves the platform, the resulting decline in SA coverage of a firm is exogenous, i.e., unrelated to contributor's expectation of how the firm's environment will change.

We define a Seeking Alpha contributor as departing if she covered at least five stocks in year $t$ and no stocks subsequently. We require that contributors cover at least 5 stocks to reduce the likelihood that the departure is related to the fundamentals of the firms being covered. Of the 5,756 Seeking Alpha contributors covering at least five stocks from 2004-2016, roughly 21% (1,201) depart Seeking Alpha. The average departing contributor writes 22 articles for 16 unique stocks in the year prior to their departure.

A firm experiences a *Contributor Departure*$_{it}$ in year $t$ if at least 20% of its existing contributors leave in year $t$.[12] We compare treated firms to candidate control firms that did not experience any departures using a difference-in-difference approach, and we require that both treated and controls firms have had coverage on Seeking Alpha for at least three years. Panel A of Table 2.7 reports summary statistics. The sample consists of 1,900 firms-years with contributor departures and 8,408 control firms-years. For both groups, we report the $\Delta$*Log* (*SA Coverage*) in year $t$, defined as *Log* (*1+SA Coverage*$_{it}$) – *Log* (*1+SA Coverage*$_{it\text{-}1}$). We observe that control firms experience a roughly 14% increase in coverage on average, whereas firms experiencing contributor departure experience a roughly 20% decline in Seeking Alpha coverage. Firms experiencing contributor departure also tend to be slightly larger and more growth-oriented than control firms.

If the choice to depart Seeking Alpha is unrelated to firm $i$'s informational environment, then we should not observe a decline in the coverage of firm $i$ by remaining Seeking Alpha contributors, sell-side analysts, or the media. To help validate the assumption that Seeking Alpha contributor departures are exogenous, we examine the relation between contributor departures and changes in total SA coverage (*SA Coverage)*, coverage by remaining SA contributors (*Non-Departing SA Coverage*), sell-side analyst coverage (*IBES Coverage*), and media coverage (*Media Coverage*), the latter defined as the total number of traditional media articles in a year.

We estimate the following panel regression:

$$\begin{aligned}\Delta(LogCoverage_{it}) = \alpha &+ \beta_1 ContributorDeparture_{it} + \beta Characteristics_{i,t-1} \\ &+ Time \times Style_{it} + \epsilon_{it},\end{aligned} \tag{2.5}$$

where $\Delta$ denotes the change from year $t$-1 to $t$, and *Characteristics* is the vector of firm characteristics included in Equation (2.1), with each variable standardized to have zero mean and unit variance. *Time*$\times$*Style*$_{it}$ is a vector of time $\times$ style indicator variables, where the style indicators are the 25 size and book-to-market groups as constructed in Daniel et al. (1997). Specifically, we first sort stocks into five quintiles based on NYSE breakpoints, and then within each size quintile, we further sort stocks into quintiles based on book-to-market. By including Time$\times$Style indicator variables, we effectively follow the approach of Kelly and Ljungqvist (2012) which matches treated firms to control firms in the same year, size quintile, and book-to-market quintile.[13] Standard errors are clustered by firm and time.

In Specification 1 in Panel B of Table 2.7, we find that firms with Seeking Alpha contributor departures experience an economically and statistically significant 35% decline in SA coverage relative to matched control firms, validating the relevance of the *Contributor Departure* instrument. On the other hand, in Specification 2 we observe that treated firms experience a significant *increase* in coverage by non-departing SA contributors, inconsistent with the idea that departures are related to firm conditions. In addition, Specifications 3-4 show no significant change in *IBES*

[12]Our results are similar if we change the cutoff for treated firms to 10%, 15%, or 25%.

[13]We find similar after including only time fixed effects.

*Coverage* or *Media Coverage* following Seeking Alpha contributor departures. The evidence in Panel B helps validate the key assumption that contributor departures are exogenous to the information environment of the firm.

### 2.6.2 Seeking Alpha Coverage and Firm Liquidity

We examine the relation between Seeking Alpha coverage and information asymmetry using the following panel regression:

$$\begin{aligned}\Delta Illiquidity_{it} = \alpha &+ \beta_1 ContributorDeparture_{it} \\ &+ \beta_2 \Delta Log(Non - DepartingSACoverage_{it}) \\ &+ \beta_3 \Delta Log(IBESCoverage_{it}) + \beta_4 \Delta Log(MediaCoverage) \\ &+ \beta Char + Time \times Style_{it} + \epsilon_{it}.\end{aligned} \tag{2.6}$$

where *Illiquidity* is the percentage bid-ask spread (*Bid-Ask*) or the Amihud (2002) illiquidity ratio (*Amihud Illiquidity*), both measured in natural logs and at the monthly frequency (by averaging all daily observations in the month).[14] We define $\Delta Illiquidity_{it}$ as the difference between $Illiquidity_{it}$

and $Illiquidity_{it\text{-}12}$. $\Delta$*Contributor Departure, Characteristics*, and *Time*$\times$*Style* are defined as in Equation (2.5). $\Delta$*Log (Non-Departing SA Coverage)*, $\Delta$ *Log (IBES Coverage)*, and $\Delta$ *Log (Media Coverage)* are proxies for changes in investor interest, defined in the Appendix. Standard errors are clustered by firm and time.

Specifications 1 and 2 of Table 2.8 report the difference-in-difference estimates for bid-ask spreads. The coefficient on *Contributor Departure* in Specification 1 indicates that relative to a portfolio of control firms matched on size, book-to-market, and year, firms with departing contributors experience a 3.45% increase in bid-ask spreads. The estimate is statistically significant at the 1% level and economically meaningful, translating into roughly 7% of the cross-sectional standard deviation of the change in bid-ask spreads (50%). Specification 2 confirms that the results are robust to controlling for changes in investor interest and firm characteristics. Specifications 3 and 4 present analogous results for the *Amihud* illiquidity measure. After including all the controls, we find a difference-in-difference of 5.15% for the *Amihud* illiquidity measure. The change in illiquidity is roughly 6% of its cross-sectional standard deviation of 87%.

We next examine the relation between contributor departures and changes in firm liquidity in event time. Specifically, we re-estimate Specifications 2 and 4 of Table 2.8, varying the timing of *Contributor Departure* from *Contributor Departure*$_{i,t\text{-}2}$ to *Contributor Departure*$_{i,t+2}$. That is, we examine the results in event time from year -2 to +2, where year 0 is the baseline Specification reported in Table 2.8. Figure 2 reports the results for *Bid-Ask* and *Amihud Illiquidity*, respectively. We find no significant changes in *Bid-Ask* or *Amihud Illiquidity* in the two years prior to the event, in support of the parallels trends assumption. As shown in Table 2.8, both liquidity measures significantly decline in year 0 (the first year after the departure of

[14] Results are similar if we aggregate illiquidity to an annual frequency.

Seeking Alpha contributors), and year 1. The two-year cumulative increase in *Bid-Ask Spread* (*Amihud Illiquidity*) is 4.84% (10.27%), which suggests that the decline in coverage has long-lived consequences for firm liquidity. Finally, we find that the change in illiquidity for both measures is small and statistically insignificant in event year 2.

*6.3 Conditioning on Firm Size, Retail Ownership, and Contributor Skill*

In this section, we explore whether the effect of Seeking Alpha coverage on liquidity depends on firm and contributor characteristics. The evidence in Table 2.6 is consistent with Seeking Alpha research being particularly valuable among smaller stocks and when the research is written by more skilled contributors. We suggest that the effect of contributor departure is more pronounced among stocks with greater retail ownership, where Seeking Alpha is likely a relatively more important information source, and when the departing contributor has greater skill.

Panel A of Table 2.9 examines whether the effect of departing contributors on liquidity varies with retail ownership by splitting the sample into two groups based on the median breakpoint of institutional ownership. Among firms with low institutional ownership, we estimate that firms with *Contributor Departures* experience a 5.26% increase in *Bid-Ask Spreads* and a 12.27% increase in the *Amihud* illiquidity measure, both of which are highly significant. In contrast, among firms in the bottom half of retail ownership, the coefficients on *Contributor Departure* for both illiquidity measures are statistically insignificant and economically small. Furthermore, the coefficient estimates across the two samples are significantly different from each other. These findings support our conjecture that the effects of Seeking Alpha on information asymmetry are much stronger among stocks heavily owned by retail investors.

Panel B splits the sample into two groups based on the median breakpoint of market capitalization. We find consistent evidence of a liquidity decline for small firms. In particular, there is a marginally significant 3.95% increase in bid-ask spreads ($p < 0.10$) and a highly significant 9.28% increase in *Amihud* illiquidity for small firms, with no evidence of a decline in liquidity for large firms.

Finally, Panel C sorts firms into two groups based on contributor skill as in Table 2.6. We find economically large and statistically significant increases in bid-ask spreads and Amihud liquidity when departing contributors have high skill. The results are weaker when departing contributors have low skill, and the differences in coefficient estimates between the two groups are significant at the 10% level. Collectively, the results in Table 2.9 confirm our conjecture that exogenous departures of Seeking Alpha contributors have a greater effect on firm liquidity for firms with greater retail ownership, smaller firms, and when a large fraction of departing contributors are highly skilled.

## 2.7 Conclusion

Individual investors are typically at an information disadvantage relative to professional investors. In recent years, innovations in technology have helped spur the democratization of investment research, with the popular provider of informative crowdsourced research Seeking Alpha playing a central role (Chen et al., 2014). In this article, we explore the extent to which crowdsourced investment research has helped level the information playing field by studying the effects of Seeking Alpha investment research on investor decision-making and the information environment of the firm.

We confirm anecdotal evidence by showing that Seeking Alpha research is geared towards retail investors, with Seeking Alpha coverage being significantly negatively related to institutional ownership and positively related to number of shareholders. We find significant increases in trading activity by retail investors on days with Seeking Alpha articles, with retail order imbalances being significantly related to the tone of research articles. More importantly, we find that Seeking Alpha research enhances the profitability of retail investor trades. In particular, the relation between retail order flow and future stock returns is roughly twice as strong on days with Seeking Alpha research articles. In contrast, we find no evidence that the informativeness of retail order flow strengthens on days when brokerage research is distributed, which is consistent with retail investors having more limited access to traditional Wall Street research. We also find no evidence that retail investors trade more profitably on days with media articles, highlighting the value of Seeking Alpha as a source of investment research.

We find that the democratization of investment research has helped improve market liquidity, which is consistent with a reduction in information asymmetry between retail investors and institutional investors. Our identification strategy relies on the idea that the departure of a Seeking Alpha contributor from the platform represents a plausibly exogenous shock to Seeking Alpha research coverage. Using a difference-in-difference approach, we find that the bid-ask spreads and the Amihud (2002) illiquidity measure increase by 2.4% and 5.5% in the year after contributor departures. These results are stronger among small firms, those with greater retail ownership, and when the departing contributors possess greater skill.

We conclude that a recent technology-induced innovation, the crowdsourcing of investment research, has helped to level the informational playing field between retail and institutional investors. We acknowledge, however, that not all innovations in information access work to level the information playing field. Many new sources of information target professional investors, and active portfolio managers expend tremendous resources to acquire investment information from Fin Tech companies (e.g. Grennan and Michaely, 2018). While Zhu (2018) finds evidence that these new sources of information help institutional investors better monitor company management, they may also work to increase information asymmetry between individuals and institutions.

Table 2.1: Summary Statistics for the Seeking Alpha Research Article Sample

The table reports information on Seeking Alpha research articles by year. The sample includes 156,513 single-ticker research articles written by 8,463 unique contributors for which the referenced stock is available in the CRSP-Compustat merged database. Average in the bottom row denotes the average across years.

| Year | Firms Covered by Seeking Alpha | Fraction of CRSP/Compustat Universe with Coverage | Seeking Alpha Articles | Seeking Alpha Contributors | Contributor -Firm Pairs | Articles per Contributor | Articles per Contributor -Firm Pair |
|---|---|---|---|---|---|---|---|
| 2005 | 240 | 5.04% | 828 | 38 | 325 | 3.45 | 1.35 |
| 2006 | 923 | 19.78% | 3,130 | 245 | 2,192 | 3.39 | 2.37 |
| 2007 | 1,437 | 31.16% | 7,368 | 561 | 4,448 | 5.13 | 3.1 |
| 2008 | 1,179 | 26.12% | 5,120 | 704 | 3,321 | 4.34 | 2.82 |
| 2009 | 1,235 | 28.95% | 7,373 | 746 | 4,176 | 5.97 | 3.38 |
| 2010 | 1,368 | 33.94% | 7,007 | 743 | 4,209 | 5.12 | 3.08 |
| 2011 | 1,338 | 34.53% | 7,093 | 945 | 4,700 | 5.3 | 3.51 |
| 2012 | 1,799 | 48.13% | 18,905 | 1,582 | 11,278 | 10.51 | 6.27 |
| 2013 | 2,322 | 64.32% | 17,550 | 1,982 | 11,683 | 7.56 | 5.03 |
| 2014 | 2,359 | 65.40% | 21,498 | 2,087 | 13,260 | 9.11 | 5.62 |
| 2015 | 2,607 | 69.76% | 22,414 | 2,059 | 13,734 | 8.6 | 5.27 |
| 2016 | 2,274 | 61.66% | 18,722 | 2,015 | 10,982 | 8.23 | 4.83 |
| 2017 | 2,217 | 62.22% | 19,505 | 1,995 | 11,545 | 8.8 | 5.21 |
| Average | 1,638 | 42.39% | 12,039 | 1,208 | 7,373 | 6.58 | 3.99 |

Table 2.2: Determinants of Research Coverage by Seeking Alpha and IBES

The table presents the results from the following panel regression:

$$Coverage_{it} = \alpha + \beta_1 Institutional\,Ownershp_{i,t-1} + \beta_2 Breadth\,of\,Ownership_{i,t-1} \\ + \beta \times Characteristicss + Time_t + \epsilon_{it}$$

In Specifications 1 and 2, Coverage is defined as the natural log of 1 plus the total number of Seeking Alpha contributors who contribute at least one article for the stock during the calendar year (SA Coverage). In Specifications 3 and 4, it is the natural log of 1 plus the total number of brokerage firms that issue at least one earnings forecast for the stock during the calendar year (IBES Coverage). Institutional $\text{Ownership}_{\text{i,t-1}}$ is the percentage of the firm's shares held by institutional investors at the end of the previous year, and Breadth of Ownership is the number of common shareholders. $\text{Char}_{\text{i,t-1}}$ is a vector of firm characteristic controls. Detailed descriptions of the variables are presented in Appendix A. The continuous variables with the exception of Return and Profitability are in natural logs, and all variables are standardized to have mean zero and unit variance. All specifications include year fixed effects and standard errors are clustered by firm, with t-statistics reported in parentheses below the corresponding coefficient estimates. The sample spans 2005-2017 and consists of 42,316 firm-year observations with 5,849 unique firm clusters.

Table 2.2, continued

| | *Log (SA Coverage)* | | *Log (IBES Coverage)* | |
|---|---|---|---|---|
| | [1] | [2] | [3] | [4] |
| Institutional Ownership | -0.18 | -0.19 | 0.15 | 0.16 |
| | (-16.27) | (-17.12) | (15.50) | (15.83) |
| Log (Breadth of Ownership) | 0.03 | 0.03 | -0.07 | -0.07 |
| | (3.13) | (3.52) | (-9.91) | (-10.05) |
| Log (Size) | 0.35 | 0.31 | 0.67 | 0.66 |
| | (20.29) | (15.85) | (49.26) | (46.84) |
| Log (BM) | -0.01 | -0.01 | 0.00 | 0.00 |
| | (-1.49) | (-1.80) | (-0.12) | (-0.05) |
| Log (Vol) | 0.14 | 0.14 | 0.05 | 0.04 |
| | (13.19) | (13.06) | (5.41) | (4.92) |
| Log (Turn) | 0.08 | 0.06 | 0.31 | 0.30 |
| | (7.75) | (5.90) | (24.40) | (24.21) |
| Return | 0.01 | 0.01 | 0.04 | 0.04 |
| | (2.44) | (2.29) | (4.30) | (4.19) |
| Profitability | 0.03 | 0.04 | -0.04 | -0.04 |
| | (5.06) | (5.45) | (-6.71) | (-6.96) |
| Log (Media Coverage) | 0.30 | 0.29 | 0.04 | 0.03 |
| | (25.46) | (25.12) | (7.26) | (5.45) |
| Log (IBES Coverage) | | 0.07 | -0.04 | |
| | | (5.32) | (-4.35) | |
| Log (SA Coverage) | | | | 0.03 |
| | | | | (5.36) |
| Fixed Effects | Time | Time | Time | Time |
| R-squared | 45.69 | 45.81 | 76.79 | 76.85 |

Table 2.3: Seeking Alpha Research Coverage and Retail Investor Trading

The table presents the results from the following daily panel regression:

$$
\begin{aligned}
RetailTrading_{it} = {} & \alpha + \beta_1 SA_{it-1,t} + \beta_2 IBES_{it-1,t} + \beta_3 Media_{t-1,t} \\
& + \beta_4 Chars_{iy-1} + Time_t + Firm_i + \epsilon_{it}
\end{aligned}
$$

In Specifications 1 and 2, Retail $\text{Trading}_{\text{it}}$ denotes Retail $\text{Turnover}_{\text{it}}$, which is 1 plus the total retail trading volume in stock i on day t, scaled by stock i's shares outstanding, measured in natural logs. In Specifications 3 and 4, it denotes Percentage Retail Turnover, which is the total retail trading volume in stock i on day t, scaled by the aggregate trading volume in stock i on day t. $\text{SA}_{\text{it-1,t}}$ is a dummy variable equal to one if Seeking Alpha issued a research report for firm i on day t or t-1. $\text{IBES}_{\text{it-1,t}}$ is a dummy variable equal to one if IBES issued a research report for firm i on day t or t-1, and $\text{Media}_{\text{it-1,t}}$ is a dummy variable equal to one if firm i was mentioned in traditional media article on day t or t-1. Char is a vector of firm characteristics measured at the end of the previous year. More details are available in the Appendix. All continuous independent variables are standardized to have mean zero and unit variance. Time and Firm indicate time (calendar day) and firm fixed effects. Standard errors are clustered by firm.

Table 2.3, continued

| | *Retail Turnover* | | *Percent Retail Turnover* | |
|---|---|---|---|---|
| | [1] | [2] | [3] | [4] |
| SA | 0.15 | 0.13 | 1.09 | 0.62 |
| | (15.07) | (24.09) | (14.77) | (20.39) |
| IBES | 0.09 | 0.08 | 0.30 | 0.06 |
| | (24.08) | (48.98) | (8.81) | (5.81) |
| Media | 0.07 | 0.09 | 0.69 | 0.12 |
| | (5.86) | (18.57) | (7.35) | (4.45) |
| Log (Size) | -0.10 | -0.12 | -2.91 | -3.79 |
| | (-7.61) | (-7.55) | (-21.09) | (-19.84) |
| Log (BM) | 0.01 | -0.03 | -0.24 | -0.16 |
| | (1.33) | (-5.11) | (-3.71) | (-2.34) |
| Inst Ownership | -0.09 | -0.01 | -2.70 | -0.95 |
| | (-16.34) | (-1.75) | (-31.19) | (-10.14) |
| Log (Breadth of Ownership) | 0.00 | 0.01 | 0.37 | 0.22 |
| | (-0.54) | (1.30) | (6.08) | (1.65) |
| Log (Vol) | 0.10 | 0.07 | 0.66 | 0.46 |
| | (12.59) | (10.82) | (6.78) | (6.15) |
| Log (Turn) | 0.17 | 0.12 | 0.70 | 0.22 |
| | (8.15) | (12.93) | (5.63) | (2.23) |
| Return | 0.01 | 0.01 | -0.54 | -0.36 |
| | (2.04) | (2.12) | (-7.59) | (-5.96) |
| Profitability | -0.03 | 0.00 | -0.70 | 0.02 |
| | (-5.93) | (0.17) | (-11.94) | (0.23) |
| Log (IBES Coverage) | 0.02 | 0.01 | -0.04 | 0.01 |
| | (1.00) | (1.56) | (-0.30) | (0.08) |
| Log (SA Coverage) | 0.06 | 0.01 | 0.75 | 0.06 |
| | (9.90) | (3.55) | (14.06) | (1.35) |
| Log (Media Coverage) | 0.03 | 0.02 | 0.50 | 0.22 |
| | (5.55) | (5.44) | (8.61) | (5.34) |
| Fixed Effects | Time | Time & Firm | Time | Time & Firm |
| Observations | 3,925,070 | 3,925,070 | 3,676,764 | 3,676,764 |
| R-squared | 27.47 | 46.13 | 35.17 | 47.33 |

Table 2.4: Seeking Alpha Article Tone and Retail Investor Order Imbalance

This table presents results from the following regression:

$$\begin{aligned} RetailOIB_{i,t+x} = \alpha &+ \beta_1 SA + \beta_2 SA \times PercentNegative_{it} \\ &+ \beta_3 SA \times PercentPositive_{it} \beta_4 SA \times ShortPosition_{it} \\ &+ \beta_5 SA \times LongPosition_{it} + \beta_6 IBES \\ &+ \beta_7 IBES \times Long + \beta_8 IBES \times Short + \beta_9 Media \\ &+ \beta_{10} Media \times PercentNegative_{it} \\ &+ \beta_{11} Media \times PercentPositive_{it} \\ &+ \beta_{12} InstOIB_{it} + \beta_{13} Char_{iy-1} + Time_t + \epsilon_{it}. \end{aligned}$$

Retail $\mathrm{OIB}_{\mathrm{i,t+x}}$ is defined as retail buy volume less retail sell volume, scaled by total retail trading volume for firm i on day t+x, where t is the event day and x varies from -3 to 3. SA is a dummy equal to one if Seeking Alpha published an article for firm i on day t; IBES and Media are defined analogously. Percent $\mathrm{Neg}_{\mathrm{it}}$ ($\mathrm{Pos}_{\mathrm{it}}$ ) is the average fraction of negative (positive) words across SA (or media) articles about company i on day t. SA Short (Long) $\mathrm{Disclosure}_{\mathrm{it}}$ equals one if the SA author is short (long) the stock, IBES Short equals one if the IBES report is negative (i.e., a recommendation downgrade or negative forecast revision), and IBES Long equals one if the IBES report is positive. Chars is a vector of firm characteristics that includes Turnover, Volatility, Size, and BM, all measured at the end of the previous year, and returns measured over the past week, month, and past two to seven months. More details are available in the Appendix. All continuous independent variables are in natural logs (with the exception of returns) and standardized to have mean zero and unit variance. Each regression includes time (calendar day) fixed effects. Standard errors are clustered by firm, and t-statistics are reported in parentheses below the corresponding coefficient estimates. The sample includes 45,084 stock-day observations over the period 2007-2017.

Table 2.4, continued

| | Retail Order Imbalance by Event Day | | | | | | |
|---|---|---|---|---|---|---|---|
| | -3 | -2 | -1 | 0 | 1 | 2 | 3 |
| SA | 2.80% | 4.35% | 5.06% | 6.52% | 5.34% | 2.39% | 3.34% |
| | (2.20) | (4.05) | (4.96) | (6.06) | (5.00) | (1.77) | (2.45) |
| SA × Percent Negative | -0.34% | -0.26% | -0.41% | -1.17% | -1.05% | -0.40% | -0.50% |
| | (-1.62) | (-1.37) | (-2.18) | (-5.52) | (-4.97) | (-1.89) | (-2.38) |
| SA × Percent Positive | 0.33% | -0.13% | 0.03% | 0.08% | 0.35% | 0.25% | -0.15% |
| | (1.29) | (-0.58) | (0.12) | (0.31) | (1.47) | (0.92) | (-0.55) |
| SA × Short Position | -0.85% | -3.08% | -4.14% | -6.75% | -6.41% | -4.33% | -2.94% |
| | (-0.49) | (-1.62) | (-2.21) | (-3.61) | (-3.41) | (-2.21) | (-1.65) |
| SA × Long Position | 1.41% | 0.59% | 1.21% | 3.83% | 2.15% | 1.52% | 1.66% |
| | (1.58) | (0.69) | (1.24) | (3.91) | (2.30) | (1.67) | (1.77) |
| Media | 1.45% | 0.94% | 2.41% | 2.14% | 2.23% | 1.18% | 0.29% |
| | (1.39) | (0.90) | (2.37) | (2.29) | (2.37) | (1.19) | (0.29) |
| Media × Percent Negative | -0.15% | -0.01% | -0.19% | -0.37% | -0.82% | -0.56% | -0.60% |
| | (-0.69) | (-0.03) | (-0.87) | (-1.84) | (-3.84) | (-2.43) | (-2.62) |
| Media × Percent Positive | -0.47% | -0.06% | -0.30% | 0.12% | 0.05% | -0.36% | 0.32% |
| | (-2.29) | (-0.32) | (-1.48) | (0.65) | (0.27) | (-1.77) | (1.55) |
| IBES | 1.51% | 2.61% | 2.27% | 1.00% | 1.65% | 1.14% | 1.67% |
| | (2.80) | (4.39) | (4.09) | (2.06) | (3.34) | (2.10) | (3.38) |
| IBES × Long | 1.83% | -1.96% | -0.29% | 2.59% | 1.63% | 0.87% | -0.62% |
| | (1.86) | (-2.04) | (-0.31) | (2.73) | (1.69) | (0.89) | (-0.65) |
| IBES × Short | 1.38% | 0.16% | 0.75% | 0.86% | -1.89% | -0.52% | -0.51% |
| | (1.44) | (0.16) | (0.74) | (0.93) | (-1.98) | (-0.54) | (-0.52) |
| Institutional OIB | 8.90% | 8.90% | 8.69% | 8.64% | 9.11% | 8.85% | 8.76% |
| | (40.20) | (38.99) | (37.99) | (38.63) | (40.32) | (39.21) | (37.67) |
| Firm Characteristics | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Fixed Effects | Time | Time | Time | Time | Time | Time | Time |
| R-squared | 2.24 | 2.24 | 2.21 | 2.19 | 2.23 | 2.25 | 2.21 |

Table 2.5: The Informativeness of Retail Trading following Seeking Alpha Research

This table reports results from the panel regression:

$$\begin{aligned}
Ret_{it,t+x} = \alpha &+ \beta_1 RetailOIB_{it} + \beta_2 InstOIB_{it} + \beta_3 SA_{it} + \beta_4 RetailOIB \times SA_{it} \\
&+ \beta_5 InstOIB_{it} \times SA_{it} + \beta_6 Media_{it} + \beta_7 RetailOIB \times Media_{it} \\
&+ \beta_8 InstOIB \times Media_{it} + \beta_9 IBES_{it} + \beta_{10} RetailOIB \times IBES_{it} \\
&+ \beta_{11} InstOIB \times IBES_{it} + \beta_{13} Char_{iy-1} + Time_t + \epsilon_{it}
\end{aligned}$$

$\mathrm{Ret}_{\mathrm{it,t+x}}$ is the return on the stock from the close of day t to the close of day t+x. Retail $\mathrm{OIB}_{\mathrm{it}}$ is the total retail buy volume less total retail sell volume, scaled by total retail trading volume for stock i on day t, and Inst OIB is defined analogously. $\mathrm{SA}_{\mathrm{it}}$ is a dummy variable equal to one if a research report on firm i was published on Seeking Alpha on day t or t-1, and IBES and Media are defined analogously. $\mathrm{RetailOIB}_{\mathrm{it}} \times \mathrm{SA}_{\mathrm{it}}$ is an interaction terms that captures the incremental informativeness of retail order imbalances following Seeking Alpha articles. Chars is a vector of firm characteristics described in the Appendix. All continuous variables are standardized and all regressions include time (calendar day) fixed effects. Standard errors are double clustered by date and firm, and t-statistics are reported below each estimate. The sample spans 2007-2017 and is comprised of 4,102,574 firm-day observations.

Table 2.5, continued

| | Holding Period | | |
|---|---|---|---|
| | [1] | [1,5] | [1,10] |
| Retail OIB | 0.02% | 0.06% | 0.09% |
| | (10.85) | (12.80) | (10.60) |
| Institutional OIB | -0.04% | -0.04% | -0.03% |
| | (-9.88) | (-4.93) | (-2.92) |
| SA | 0.02% | 0.04% | 0.04% |
| | (2.26) | (1.28) | (0.71) |
| Retail OIB × SA | 0.04% | 0.04% | 0.09% |
| | (4.19) | (1.49) | (2.58) |
| Institutional OIB × SA | 0.02% | 0.02% | -0.01% |
| | (2.02) | (0.62) | (-0.13) |
| Media | -0.01% | -0.01% | -0.01% |
| | (-1.38) | (-0.29) | (-0.15) |
| Retail OIB × Media | 0.02% | 0.04% | 0.03% |
| | (1.24) | (1.24) | (0.85) |
| Institutional OIB × Media | 0.07% | 0.11% | 0.09% |
| | (4.31) | (3.21) | (2.40) |
| IBES | -0.01% | 0.00% | 0.01% |
| | (-1.05) | (-0.08) | (0.12) |
| Retail OIB × IBES | 0.00% | 0.00% | -0.02% |
| | (-0.24) | (-0.31) | (-0.90) |
| Institutional OIB × IBES | 0.02% | 0.01% | 0.04% |
| | (2.25) | (0.54) | (1.96) |
| $Ret_{w-1}$ | -0.02% | -0.09% | -0.12% |
| | (-4.01) | (-3.77) | (-3.31) |
| $Ret_{m-1}$ | -0.01% | -0.04% | -0.04% |
| | (-1.87) | (-1.23) | (-0.72) |
| $Ret_{m-7,m-2}$ | 0.00% | 0.00% | 0.00% |
| | (-0.50) | (-0.10) | (0.04) |
| $Turnover_{m-1}$ | -0.01% | -0.06% | -0.12% |
| | (-2.37) | (-2.27) | (-2.52) |
| $Volatility_{m-1}$ | 0.01% | 0.06% | 0.09% |
| | (1.84) | (1.50) | (1.22) |
| Log (Size) | 0.00% | -0.01% | -0.01% |
| | (-0.62) | (-0.32) | (-0.24) |
| Log (BM) | 0.00% | 0.01% | 0.03% |
| | (0.39) | (0.50) | (0.52) |
| Fixed Effects | Time | Time | Time |
| R-squared | 16.59% | 15.81% | 15.24% |

Table 2.6: Retail Trading Informativeness Following Seeking Alpha Research: Size and Contributor Skill

The table repeats the analysis in Table 5 after partitioning the sample based on firm size and contributor skill. In Panel A, the sample is split into Small and Large firms based on median market capitalization at the end of the previous year. In Panel B, the research article sample is split into High and Low Skill contributors based on the median average two-day absolute market-adjusted return across the last five articles written by the contributor. Standard errors are double clustered by date and firm, and t-statistics are reported in parentheses below each estimate. The sample spans 2007-2017

| | Coefficient on Retail OIB×SA | | |
|---|---|---|---|
| Panel A: Firm Size | | | |
| | Holding Period: | | |
| | [1] | [1,5] | [1,10] |
| Small Firms | 0.09% | 0.10% | 0.17% |
| | (4.49) | (2.33) | (2.78) |
| Large Firms | 0.01% | -0.01% | 0.04% |
| | (0.56) | (-0.42) | (1.08) |
| Difference in Coefficients | 0.08% | 0.12% | 0.14% |
| | (3.62) | (2.10) | (1.95) |
| | | | |
| Panel B: Contributor Skill | | | |
| | Holding Period: | | |
| | [1] | [1,5] | [1,10] |
| High Skill | 0.11% | 0.14% | 0.25% |
| | (4.22) | (2.70) | (2.77) |
| Low Skill | 0.01% | -0.08% | -0.07% |
| | (0.71) | (-1.87) | (-1.24) |
| Difference in Coefficients | 0.09% | 0.22% | 0.32% |
| | (3.05) | (3.33) | (2.91) |

Table 2.7: Departing Seeking Alpha Contributors: Validity Tests

The table reports the effects of departing Seeking Alpha contributors on research coverage. Firms experience Contributor Departure if at least 20% of the firm's existing contributors depart Seeking Alpha in the previous year. Control firms are firms with no departing Seeking Alpha contributors. Panel A reports univariate statistics for coverage and firm characteristics. SA Coverage is the total number of unique Seeking Alpha contributors writing at least one article for the stock during the calendar year, and $\Delta$Log (SA Coverage) is defined as Log(1+SA $\text{Coverage}_{\text{it}}$)– Log(1+SA $\text{Coverage}_{\text{it-1}}$). Panel B presents the estimates from regressing measures of changes in different measures of research Coverage on a Contributor Departure indicator variable. We include as controls a number of firm characteristics described in the appendix. The regressions also include time $\times$ style fixed effects, where the style dummies capture the 25 size and book-to-market portfolios, which effectively matches treated firms to control firms in the same year and size and book-to-market quintiles. Standard errors are clustered by firm and time, and t-statistics are reported in parentheses below each estimate. The sample covers 2005-2017.

Table 2.7, continued

| | Coefficient on Retail OIB×SA | | | |
|---|---|---|---|---|
| Panel A: Summary Statistics for Control Firms and Contributor Departure Firms | | | | |
| | Observations | Percent Change Coverage | Log (Size) | Log (MB) |
| Control Firms | 8,408 | 14.06% | 13.60 | 0.78 |
| Contributor Departure Firms | 1,900 | -19.61% | 14.50 | 1.07 |

Panel B: Regression of Changes in Coverage on Contributor Departure

| | Δ Log SA Coverage | Δ (Log Non -Depart SA Coverage) | Δ (Log IBES Coverage) | Δ (Log Media Coverage) |
|---|---|---|---|---|
| | [1] | [2] | [3] | [4] |
| Contributor Departure | -35.40% | 6.31% | -0.22% | 4.02% |
| | (-16.15) | (2.60) | (-0.59) | (1.33) |
| Δ (Log Non-Depart. SA Coverage) | | | 1.44% | 4.94% |
| | | | (6.57) | (3.79) |
| Δ (Log IBES Coverage ) | 9.50% | 9.63% | | (0.01) |
| | (5.70) | (5.70) | | (0.36) |
| Δ (Log Media Coverage ) | 11.18% | 10.85% | 0.22% | |
| | (7.69) | (7.30) | (0.36) | |
| Log (Size) | 1.90% | 0.57% | 9.85% | 12.38% |
| | (0.74) | (0.21) | (6.03) | (3.13) |
| Log (BM) | -3.10% | -3.09% | -0.21% | -0.58% |
| | (-2.22) | (-2.19) | (-0.20) | (-0.77) |
| Inst Ownership | 0.18% | 0.78% | 1.40% | -1.56% |
| | (0.12) | (0.49) | (3.43) | (-1.50) |
| Log (Breadth of Ownership) | 0.65% | 0.67% | -1.01% | 1.15% |
| | (0.97) | (0.92) | (-4.93) | (1.48) |
| Log (Vol) | 1.09% | 0.77% | 1.74% | 4.44% |
| | (0.65) | (0.46) | (2.61) | (2.48) |
| Log (Turn) | -2.54% | -3.04% | 3.98% | -1.03% |
| | (-1.82) | (-1.96) | (2.66) | (-1.68) |
| Return | 1.78% | 1.79% | 1.82% | 0.11% |
| | (2.01) | (1.95) | (5.41) | (0.21) |
| Profitability | 1.57% | 1.51% | 0.19% | 0.28% |
| | (1.47) | (1.39) | (0.45) | (0.41) |
| Log (IBES Coverage) | 2.93% | 3.08% | -16.65% | 1.43% |
| | (3.03) | (3.74) | (-9.73) | (1.06) |
| Log (Media Coverage) | 4.52% | 3.66% | 0.07% | -21.58% |
| | (3.58) | (3.19) | (0.13) | (-5.08) |
| Fixed Effects | $Time \times Style$ | $Time \times Style$ | $Time \times Style$ | $Time \times Style$ |
| Total R-squared | 16.01% | 12.23% | 18.21% | 29.30% |

Table 2.8: Departing Seeking Alpha Contributors: Effects on Firm Liquidity

The table reports the results from the following panel regression:

$$\begin{aligned}\Delta Illiquidity_{it} = \alpha &+ \beta_1 ContributorDeparture_{it} \\ &+ \beta_2 \Delta Log(Non - DepartingSACoverage_{it}) \\ &+ \beta_3 \Delta Log(IBESCoverage_{it}) + \beta_4 \Delta Log(MediaCoverage) \\ &+ \beta Char + Time \times Style_{it} + \epsilon_{it}.\end{aligned}$$

where $\mathrm{Illiquid_{it}}$ is the percentage bid-ask spread (Bid-Ask) in Specifications 1 and 2 and the Amihud (2002) illiquidity ratio (Amihud) in Specifications 3 and 4, both measured in natural logs. Both measures are calculated monthly using daily averages; change is defined as $\mathrm{Illiquid_{it}}$ - $\mathrm{Illiquid_{it\text{-}12}}$. Firm i experiences contributor departure ($\mathrm{ContrDepart_{it}}$) if at least 20% of its existing contributors depart Seeking Alpha (i.e. issue research for at least five stocks in the calendar year prior to month t, and for no stocks in calendar year 0 onwards). Non-Depart SA $\mathrm{Cov_{it}}$ ($\mathrm{IBESCov_{it}}$) denotes the number of unique non-departing Seeking Alpha contributors (brokerage firms in IBES) that issue at least one research report research for the stock during calendar year t. Chars denotes a vector of firm characteristics that is defined in the Appendix. Each regression includes time×style fixed effects ($\mathrm{Tm{\times}Style_{it}}$) based on the 25 size and book-to-market portfolios. Standard errors are clustered by firm and month, and t-statistics are reported below each estimate. The sample period spans 2005-2017 and is comprised of 123,645 firm-month observations.

Table 2.8, continued

| | Δ Log (Bid-Ask) | | Δ Log (Amihud) | |
|---|---|---|---|---|
| | [1] | [2] | [3] | [4] |
| *Contributor Departure* | 3.45% | 2.15% | 7.51% | 5.15% |
| | (2.98) | (2.47) | (7.02) | (5.72) |
| *Log (Size)* | | -7.82% | | -16.17% |
| | | (-1.87) | | (-3.69) |
| *Log (BM)* | | 1.28% | | 1.86% |
| | | (1.57) | | (1.09) |
| *Institutional Ownership* | | -0.42% | | -4.71% |
| | | (-0.46) | | (-3.07) |
| *Log (Breadth of Ownership)* | | 0.04% | | -0.16% |
| | | (0.12) | | (-0.28) |
| *Log (Vol)* | | -5.68% | | -16.09% |
| | | (-2.79) | | (-4.08) |
| *Log (Turnover)* | | 6.06% | | 20.95% |
| | | (6.57) | | (10.00) |
| *Return* | | -13.95% | | -27.62% |
| | | (-10.57) | | (-10.56) |
| *Profitability* | | -0.45% | | -1.74% |
| | | (-0.69) | | (-2.17) |
| *Log (IBES Coverage)* | | 3.54% | | -0.38% |
| | | (2.59) | | (-0.25) |
| Δ *(Log Non-Departing SA Coverage )* | | 0.54% | | 0.63% |
| | | (0.72) | | (0.80) |
| Δ *(Log IBES Coverage )* | | -1.95% | | -7.73% |
| | | (-3.43) | | (-7.05) |
| Δ *(Log Media Coverage )* | | -11.77% | | -28.02% |
| | | (-3.91) | | (-6.48) |
| Fixed Effects | $Time \times Style$ | $Time \times Style$ | $Time \times Style$ | $Time \times Style$ |
| R-squared | 19.92% | 27.86% | 22.54% | 34.60% |

Table 2.9: Departing Seeking Alpha Contributors and Liquidity: Ownership, Size, and Contributor Skill

The table repeats the analysis in Table 8 after partitioning the sample based on institutional ownership, firm size, and contributor skill. In Panel A (B), the sample of firms is split into High and Low Institutional Ownership (Size) using the median value at the end of the previous year. In Panel C, the sample split is based on whether the fraction of departing contributors with high skill is greater than 50%. Contributor skill is measured as the average two-day market-adjusted return across the last five articles written by the contributor. For each partition, we estimate the panel regression:

$$\begin{aligned}\Delta Illiquidity_{it} = &\alpha + \beta_1 ContributorDeparture_{it}\\ &+ \beta_2 \Delta Log(Non - DepartingSACoverage_{it})\\ &+ \beta_3 \Delta Log(IBESCoverage_{it}) + \beta_4 \Delta Log(MediaCoverage)\\ &+ \beta Chars + Time \times Style_{it} + \epsilon_{it}.\end{aligned}$$

where $Illiquid_{it}$ is the percentage bid-ask spread (*Bid-Ask*) or the Amihud (2002) illiquidity ratio (*Amihud*), both measured in natural logs. The illiquidity measures are calculated monthly using daily averages, and $\Delta Illiquid_{it}$ is the illiquidity measure in month $t$ less the measure in month $t$-12. Firms are classified as experiencing contributor departure ($ContrDepart_{it}$) if at least 20% of firm $i$'s existing contributors depart Seeking Alpha (i.e. issue research for at least five stocks in calendar years -3 through -1 before month $t$, and for no stocks in calendar year 0 onwards). $Non\text{-}DepartSACov_{it}$ ($IBESCov_{it}$) denotes the number of unique non-departing Seeking Alpha contributors (brokerage firms in IBES) that issue at least one research report research for stock during calendar year 0. *Chars* denotes a vector of firm characteristics that is defined in the Appendix. Each regression includes time×style fixed effects ($Tm \times Style_{\mathrm{it}}$) based on the 25 size and book-to-market portfolios. Standard errors are clustered by firm and time, and $t$-statistics are reported below each estimate. The sample period spans 2005-2017 and is comprised of 123,645 firm-month observations.

Table 2.9, continued

| | Coefficient on Contributor Departure | |
|---|---|---|
| **Panel A: Sorts on Institutional Ownership** | | |
| | Δ *Log (Bid-Ask)* | Δ *Log (Amihud)* |
| Low Institutional Ownership | 5.26 | 12.27 |
| | (3.08) | (5.40) |
| High Institutional Ownership | -0.06 | -0.25 |
| | (-0.07) | (-0.27) |
| Difference in Coefficients | 5.32 | 12.52 |
| | (2.68) | (4.35) |

| | | |
|---|---|---|
| **Panel B: Sorts on Firm Size** | | |
| | Δ *Log (Bid-Ask)* | Δ *Log (Amihud)* |
| Small Firms | 3.95 | 9.28 |
| | (1.67) | (4.99) |
| Large Firm | 0.60 | 0.95 |
| | (0.46) | (0.94) |
| Difference in Coefficients | 3.35 | 8.33 |
| | (1.04) | (3.47) |

| | | |
|---|---|---|
| **Panel C: Sorts on Departing Contributor Skill** | | |
| | Δ *Log (Bid-Ask)* | Δ *Log (Amihud)* |
| High Skill | 3.95 | 7.25 |
| | (3.54) | (5.21) |
| Low Skill | 0.61 | 3.43 |
| | (0.66) | (3.19) |
| Difference in Coefficients | 3.34 | 3.82 |
| | (2.26) | (1.90) |

## Appendix A: Variable Definitions for Chapter 1

- *Fitness* – The fitness represents how closely the modeled portfolio replicates the daily returns of the actual fund. It is the result from minimizing equation 1.18:

$$\min_{k_1 \ldots k_N, \%Cash} \sum_{t=1}^{T} \left[ \left( \frac{R_{t,F,reported}}{(1 - \%Cash)} - \sum_{i=1}^{N} R_{i,t} \times \frac{S_{i,k,t-1} \times P_{i,t-1}}{\sum_{i=1}^{N} S_{i,k,t-1} \times P_{i,t-1}} \right)^2 + 0.01V_t \right].$$

- $flow_t$ – The net flows to the mutual fund in quarter t, scaled by total net assets.
- $Mom1\&5$ – For each stock, I take the average returns over the past 1 and 5 days. For each fund quarter, I calculate the principle-weighted average $Mom1\&5$ of stocks purchased minus stocks sold, with respect to the date on which each transaction is made. This measure is computed only within the Ancerno sample.
- $Mom1\&5_{calc}$ – This measure is analogous to $Mom1\&5$, except that the dates on which the trades take place are first estimated using the fund's quarterly holdings and daily returns.
- $Mom1\&5_{end}$ – This measure is analogous to $Mom1\&5$, except that the dates on which the trades are assumed to take place on the final day of the quarter.
- $Mom1\&5Q$ – This measure is formed by taking each stock's average $Mom1\&5$ throughout the quarter and computing the principle-weighted average of stocks purchased minus stocks sold.
- $Mom1\&5_{smooth}$ – This measure is analogous to $Mom1\&5$, except that the quarterly trades are assumed to take place in equal quantities each day within the quater.

## Appendix B: Variable Definitions for Chapter 2

*B.1 Seeking Alpha Variables*

- *SA Coverage* – the number of unique Seeking Alpha contributors writing an opinion article for a firm during the calendar year (Source: Seeking Alpha).
- *SA Articles* – the total number of Seeking Alpha opinion articles written for a firm during the calendar year (Source: Seeking Alpha).
- *Contributor Departure* – an indicator variable equal to one if a firm has lost 20% or more of its existing coverage due to *plausibly exogenous departures* from Seeking Alpha. (Source: Seeking Alpha).
  - We consider dropped coverage for firm $i$ in year $t$ to be plausibly exogenous if 1) the departing contributor was covering at least five firms (including firm $i$) in year *t-1* and 2) the contributor never issues research for any firms on *Seeking Alpha* at any point after year *t-1*.
- *Non-Departing SA Coverage* – defined as *SA Coverage* less the total number of contributors who dropped coverage due to plausible exogenous departures (as defined above).
- *SA* – a dummy variable equal to one if a Seeking Alpha opinion article was written about firm $i$ on day $t$ or day *t-1* (Source: Seeking Alpha).
- *Percent Negative (Positive)* – the average fraction of negative (positive) words across all single-ticker articles published on Seeking Alpha about firm $i$ on day $t$. (Source: Seeking Alpha). The list of negative and positive words is taken from Loughran and McDonald (2011).
- *Short (Long) Position* – a dummy variable equal to one if the author discloses a short (long) position about the company discussed in the article. This measure is average across all single-ticker articles published about firm $i$ on day $t$.
- *Contributor Skill* – the two day absolute market-adjusted return averaged across the past five articles written by the contributor. (Source: Seeking Alpha/CRSP).

*B.2 Liquidity Measures*

- *Retail Turnover* – average daily retail turnover (i.e., retail share volume scaled by shares outstanding) during the calendar year. Retail trading is classified using the approach outlined in Boehmer, Jones, and Zhang (2017). (Source: TAQ and CRSP).
- *Percent Retail Turnover* – retail share volume scaled by total share volume. Retail trading is classified using the approach outlined in Boehmer, Jones, and Zhang (2017). (Source: TAQ and CRSP).

- *Amihud* - the Amihud (2002) illiquidity measure computed using all daily data available in the calendar year.

- *Bid-Ask Spread* – the average daily bid-ask spread computed as the difference between the (end of day) bid and ask price, divided by the midpoint. Winsorized at the 1st and 99th percentiles. (Source: CRSP).

*B.3 Other Variables:*

- *Size* – the market capitalization computed as share prices times total shares outstanding at the end of the year (Source: CRSP).

- *Book-to-Market (BM)* – the book-to-market ratio computed as the book value of equity during the calendar year scaled by the market capitalization at the end of the calendar year. Negative values are deleted and positive values are winsorized at the 1st and 99th percentile. (Source: CRSP/Compustat).

- *Volatility* – the standard deviation of daily returns during the calendar year (Source: CRSP).

- *Age* – the number of years since the Initial Public Offering (Source: Compustat).

- *Profitability* – EBITDA scaled by book value of assets. Winsorized at the 1st and 99th percentiles (Source: Compustat).

- $Return_{t\text{-}1,\, t\text{-}12}$ – the buy-and-hold gross return over the prior 12 months. Alternative holding periods are labelled analogously (Source: CRSP).

- *Institutional Ownership* – the percentage of the firm's shares held by institutions at year end (Source: Thomson Reuters Institutional Holdings S34).

- *Retail Ownership* – 1 – *Institutional Ownership*.

- *Breadth of Ownership* – the total number of common shareholders (Source: Compustat).

- *IBES Coverage* – the number of unique brokerage houses issuing an earnings forecast for a firm during the calendar year (Source: IBES).

- *Media Coverage* – the total number of media articles about a firm during the calendar year (Source: Factiva and Chen et al., 2014)).

- *IBES* – an indicator variable equal to one if an IBES earnings forecast or IBES investment recommendation was issued for a firm on day *t* or day *t-1* (Source: IBES).

- *Media* – an indicator variable equal to one if a Media article was issued for a firm on day *t* or day *t-1*. (Source: Factiva and Chen et al., 2014).

- *Earnings Event* – an indicator variable equal to one if earnings or earnings guidance is announced for the firm for day *t* or day *t-1* (Source: IBES).

- *Retail OIB* – retail buy volume less retail sell volume, scaled by total retail trading volume. Retail trading is classified using the approach outlined in Boehmer, Jones, and Zhang (2017), and trades are signed using the Lee and Ready (1991) algorithm (Source: TAQ).

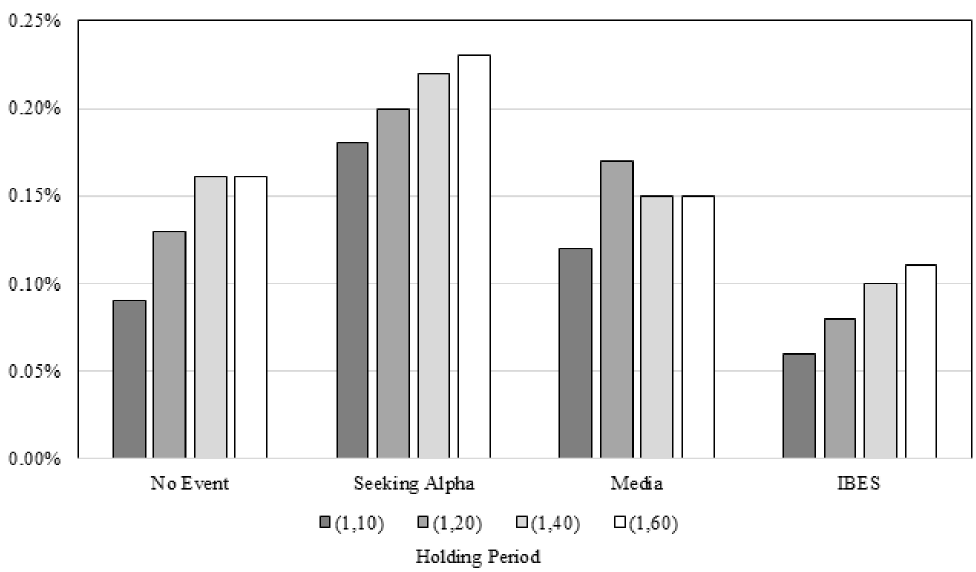
- *Institutional OIB* – the total (non-retail) share volume bought less the (non-retail) share volume sold, scaled by the total (non-retail) volume traded. Retail trading is classified using the approach outlined in Boehmer, Jones, and Zhang (2017), and trades are signed using the Lee and Ready (1991) algorithm (Source: TAQ).

Figure B1: The Informativeness of Retail Trading Following Seeking Alpha Research over Time

Each month, we estimate the following panel regression:

$$
\begin{aligned}
Ret_{it,t+10} = {} & \alpha + \beta_1 RetailOIB_{it} + \beta_2 InstOIB_{it} + \beta_3 SA_{it} + \beta_4 RetailOIB \times SA_{it} \\
& + \beta_5 InstOIB_{it} \times SA_{it} + \beta_6 Media_{it} + \beta_7 RetailOIB \times Media_{it} \\
& + \beta_8 InstOIB \times Media_{it} + \beta_9 IBES_{it} + \beta_{10} RetailOIB \times IBES_{it} \\
& + \beta_{11} InstOIB \times IBES_{it} + \beta_{13} Char_{iy-1} + Time_t + \epsilon_{it}
\end{aligned}
\tag{7}
$$

The regression is identical to Specification 3 of Table 5 except that the regression is estimated separately each month. The figure plots the cumulative coefficient on *Retail OIB * SA* over the full-sample period.



## References


Asthana, Balsam S., S., and S. Sankaraguruswamy, 2004, Differential response of small versus large investors to 10-k filings on edgar., *The Accounting Review* 79, 571–589.

Avery, Chevalier J., C., and R. Zeckhauser, 2016, The "caps" prediction system and stock market returns., *Review of Finance* 20, 1363–1381.

Barber, B., and T. Odean, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors., *Journal of Finance* 55, 773–806.

Barber, B., and T. Odean, 2002, Online investors: Do the slow die first?, *Review of Financial Studies* 15, 455–488.

Barber, B., and T. Odean, 2013, The behavior of individual investors., *Handbook of the Economics of Finance, Elsevier* 2, 1533–1570.

Barber, Odean T., B., and N. Zhu, 2009, Do retail trades move markets?, *Review of Financial Studies* 22, 151–186.

Barclay, Michael J, and Jerold B Warner, 1993, Stealth trading and volatility: Which trades move prices?, *Journal of Financial Economics* 34, 281–305.

Battalio, Corwin S., R., and R. Jennings, 2016, Can brokers have it all? on the relation between make-take fees and limit order execution quality., *Journal of Finance* 71, 2913–2238.

Bergstra, James, and Yoshua Bengio, 2012, Random search for hyper-parameter optimization, *Journal of Machine Learning Research* 13, 281–305.

Bhattacharya, Cho Y., N., and J. Kim, 2018, Levelling the playing field between large and small institutions: evidence from the sec's xbrl mandate., *Accounting Review* forthcoming.

Bhushan, R., 1989, Firm characteristics and analyst following., *Journal of Accounting and Economics* 11, 255–274.

Blankespoor, Miller B., E., and H. White, 2014, Initial evidence on the market impact of the xbrl mandate., *Review of Accounting Studies* 19, 1468–1503.

Boehmer, Jones C., E., and X. Zhang, 2017, Tracking retail investor activity, Working paper.

Brown, Call A. Clement M., L., and N. Sharp, 2015, Inside the "black box" of sell-side financial analysts., *Journal of Accounting Research* 53, Jan–47.

Campbell, DeAngelis M., J., and J. Moon, 2016, Skin in the game: Personal stock holdings and investors' response to stock analysis on social media, Working paper.

Campbell, John Y, Tarun Ramadorai, and Allie Schwartz, 2009, Caught on tape: Institutional trading, stock returns, and earnings announcements, *Journal of Financial Economics* 92, 66–91.

Carhart, Mark M, 1997, On persistence in mutual fund performance, *The Journal of Finance* 52, 57–82.

Chen, De P. Hu J., H., and B.H. Hwang, 2014, Wisdom of the crowds: The value of stock opinions transmitted through social media., *Review of Financial Studies* 27, 1367–1403.

Chordia, Tarun, 1996, The structure of mutual fund charges, *Journal of Financial Economics* 41, 3–39.

Cohen, Lauren, Dong Lou, and Christopher J Malloy, 2016, Cloaked trading, *Journal of Investment Consulting* 17, 69–80.

Corrie Driebusch, Gregory Zuckerman, Alexander Osipovich, 2018, What's the biggest trade on the new york stock exchange? the last one, *Wall Street Journal, March 14* .

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

Daniel, Grinblatt M. Titman S., K., and R. Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks., *Journal of Finance* 52, 1035–1058.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *The Journal of finance* 52, 1035–1058.

Del Guercio, Diane, and Paula A Tkac, 2002, The determinants of the flow of funds of managed portfolios: Mutual funds vs. pension funds, *Journal of Financial and Quantitative Analysis* 37, 523–557.

Duarte, Han X. Harford J., J., and L. Young, 2008, Information asymmetry, information dissemination, and the effect of regulation fd, *Journal of Financial Economics* 87, 24–44.

Eleswarapu, Thomspon R., V., and K. Venkataraman, 2004, The impact of regulation fair disclosure: Trading costs and information asymmetry., *Journal of Financial and Quantitative Analysis* 39, 209–225.

Fama, Eugene F, and Kenneth R French, 2010, Luck versus skill in the cross-section of mutual fund returns, *The journal of finance* 65, 1915–1947.

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of political economy* 81, 607–636.

Franzoni, Francesco, and Alberto Plazzi, 2013, Do hedge funds provide liquidity? evidence from their trades, Working paper. University of Lugano and Swiss Finance Institute.

Frazzini, A., and O. Lamont, 2008, Dumb money: Mutual fund flows and the cross-section of stock returns., *Journal of Financial Economics* 88, 299–322.

Green, Jame R. Markov S., T.C., and M. Subasi, 2014, Broker-hosted investor conferences., *Journal of Accounting and Economics* 58, 142–166.

Grennan, J., and R. Michaely, 2018, Fintechs and the market for financial analysis, Working paper.

Ha, JinGi, and Jianfeng Hu, 2017, How smart is institutional trading?, Working Paper.

Huang, J., and M. Gao, 2018, Informing the market: the effect of modern information technologies on information production, Working paper.

Hvidkjaer, S., 2008, Small trades and the cross-section of stock returns., *Review of Financial Studies* 21, 1123–1151.

Ivković, Zoran, and Scott Weisbenner, 2009, Individual investor mutual fund flows, *Journal of Financial Economics* 92, 223–237.

Jame, Johnston R. Markov S., R., and M. Wolfe, 2016, The value of crowdsourced earnings forecasts., *Journal of Accounting Research* 54, 1077–1110.

Jame, Russell, 2018, Liquidity provision and the cross section of hedge fund returns, *Management Science* 64, 3288.

Jordan, Bradford, and Timothy Riley, 2016, Skill and persistence in mutual fund returns: Evidence from a six-factor model, Working Paper.

Jylhä, Petri, Kalle Rinne, and Matti Suominen, 2014, Do hedge funds supply or demand liquidity?, *Review of Finance* 18, 1259–1298.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, 2006, Unobserved actions of mutual funds, *The Review of Financial Studies* 21, 2379–2416.

Kaniel, Liu S. Saar G., R., and S. Titman, 2012, Individual investor trading and return patterns around earnings announcements, *Journal of Finance* 67, 639–680.

Kaniel, Saar G., R., and S. Titman, 2008, Individual investor sentiment and stock returns., *Journal of Finance* 63, 273–310.

Kelley, E., and P. Tetlock, 2013, How wise are crowds? insights from retail orders and stocks returns., *Journal of Finance* 68, 1229–1265.

Kelley, E., and P. Tetlock, 2017, Retail short-selling and stock prices., *Review of Financial Studies* 30, 801–834.

Kelly, B., and A. Ljungqvist, 2012, Testing asymmetric-information asset pricing models., *Review of Financial Studies* 25, 1366–1413.

Kumar, A., and C. Lee, 2006, Retail investor sentiment and return comovements., *Journal of Finance* 61, 2451–2486.

Loughran, T., and B. McDonald, 2011, When is a liability not a liability? textual analysis, dictionaries, and 10-ks., *Journal of Finance* 66, 35–65.

McCall, John, 2004, Genetic algorithms for modelling and optimisation, *Journal of Computational and Applied Mathematics* 184, 205–222.

Nagel, Stefan, 2012, Evaporating liquidity, *The Review of Financial Studies* 25, 2005–2039.

Newey, Whitney K, and Kenneth D West, 1987, Hypothesis testing with efficient method of moments estimation, *International Economic Review* 777–787.

Petersen, Mitchell A, 2009, Estimating standard errors in finance panel data sets: Comparing approaches, *The Review of Financial Studies* 22, 435–480.

Puckett, Andy, and Xuemin Sterling Yan, 2011, The interim trading skills of institutional investors, *The Journal of Finance* 66, 601–633.

Rinne, Kalle, and Matti Suominen, 2014, Mutual funds' returns from providing liquidity and costs of immediacy, Working Paper.

Seru, A., T. Shumway, and N. Stoffman, 2009, Learning by trading., *Review of Financial Studies* 23, 705–739.

Sharpe, William F., 1992, Asset allocation: Management style and performance measurement, *The Journal of Portfolio Management* 18, 7–19.

Sirri, Erik R, and Peter Tufano, 1998, Costly search and mutual fund flows, *The journal of finance* 53, 1589–1622.

Spiess, Andrej-Nikolai, and Natalie Neumeyer, 2010, An evaluation of r2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a monte carlo approach, *BMC Pharmacology* 10, 6.

Tetlock, P., 2007, Giving content to investment sentiment: The role of media in the stock market., *Journal of Finance* 62, 1139–1168.

White, Halbert, 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica: Journal of the Econometric Society* 817–838.

Yan, Xuemin, 2006, The determinants and implications of mutual fund cash holdings: Theory and evidence, *Financial Management* 35, 67–91.

Yeniay, Özgür, 2005, Penalty function methods for constrained optimization with genetic algorithms, *Mathematical and computational Applications* 10, 45–56.

Zheng, Lu, 1999, Is money smart? a study of mutual fund investors' fund selection ability, *the Journal of Finance* 54, 901–933.

Zhu, C., 2018, Big data as a governance mechanism, *Review of Financial Studies* forthcoming.

**Vita**

# Michael Farrell

## EDUCATION

**Concordia University**
MSc. Finance 2014

**McGill University**
B.A. Economics 2009

## WORKING PAPERS

"Read Between the Filings: Daily Mutual Fund Holdings and Liquidity Provision"

- Presented at FMA Doctoral Consortium (2018), EFA (2019), Sacred Heart University, Elon University, SUNY Albany, Saint Joseph's University, Seton Hall University, and Manhattan College.

"The Democratization of Investment Research: Implications for Retail Investor Profitability and Firm Liquidity" (with T. Clifton Green, Russell Jame, and Stanimir Markov)

- Presented at the Hawaii Accounting Research Conference (HARC 2019),* Financial Accounting and Reporting Section Midyear Meeting (FARS 2019),* Smokey Mountain Finance Conference (2019),* Midwest Finance Association (MFA 2019), Western Finance Association (WFA 2019)*, and SFS Cavalcade (2019).*

(* indicates presentations by coauthor)

## TEACHING

**University of Kentucky**
Introduction to Corporate Finance
Summer 2017. *Average Rating 4.85/5.0*
Investment Analysis
Summer 2016. *Average Rating 4.54/5.0*

## FELLOWSHIPS AND AWARDS

### University of Kentucky

| | |
|---|---|
| 2017-2019 | Gatton Fellowship |
| 2014-2017 | Luckett Fellowship |
| 2016 | Outstanding Graduate Assistant Award |
| 2014–2015 | Max Steckler Award |
| 2013 | MSc Best Research Presentation |
| 2012 | Scotiabank M.Sc. Fellowship in Finance |

### Professional Activities

| | |
|---|---|
| 2019 | Discussant and Session Chair at EFA Conference |
| 2019 | Referee for Pacific Basin Journal of Finance |
| 2019 | Discussant at MFA Conference |
| 2018 | FMA Doctoral Consortium Participant |
| 2018 | Referee for Managerial Finance |
| 2016 | Session Chair FMA Conference |
| 2015-2016 | Discussant at FMA Conference |