# Read Between the Filings
## Daily Mutual Fund Holdings and Liquidity Provision

Michael Farrell *

August 2018

**ABSTRACT**

Many questions about mutual fund trading require daily holdings, yet mutual funds are only required to report quarterly holdings. I model intraquarter trading and use the 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 naive 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.

*PhD candidate at the Gatton College of Business and Economics, University of Kentucky. Email: michael.farrell@uky.edu. I thank Russell Jame, Adam Krzyzak, and brown bag participants at the University of Kentucky for helpful insights.

# 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 et al. (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 nonroundtrip 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

[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.

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 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

[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.

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.

# 2 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 et al., 1997). On the other hand, meaningful activity takes place in between filings, producing a persistent "return gap" Kacperczyk et al. (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.

## 2.1 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, mutual funds typically do not exhibit trading skill (Fama and French, 2010). As a result, the incremental effect of round trip trades on portfolio returns can reasonably be assumed to be noise. In addition, 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:

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

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:

$$w_{it} = \frac{S_{i,t-1} \times P_{i,t-1}}{\sum_{i=1}^{N} S_{i,t-1} \times P_{i,t-1}}, \tag{2}$$

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:

$$S_{i,t} = S_{i,0} + \sum_{d=1}^{t} \Delta S_{i,d} \tag{3}$$

Combining equations (1), (2), and (3), the fund's return on a given day can be approximated by:

$$R_{t,F} \approx \sum_{i=1}^{N} R_{i,t} \times \frac{\left(S_{i,0} + \sum_{d=1}^{t-1} \Delta S_{i,d}\right) \times P_{i,t-1}}{\sum_{i=1}^{N} \left(S_{i,0} + \sum_{d=1}^{t-1} \Delta S_{i,d} \times P_{i,t-1}\right)} = R_{t,M}. \tag{4}$$

$R_{t,M}$ 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, $\Delta S_i$:

$$\Delta S_{i,d} \in \{1, 2, 3, \ldots, \Delta S_i\}, \tag{5}$$

and

$$\Delta S_i = S_{i,T} - S_{i,0}, \tag{6}$$

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:

$$\Delta S_i = \sum_{d=1}^{T} \Delta S_{i,d} \tag{7}$$

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:

$$|\Delta S_i| = \sum_{d=1}^{T} |\Delta S_{i,d}|. \tag{8}$$

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 $\Delta S_{i,d}$:

$$\min_{S_{1,1} \ldots S_{N,T}} \sum_{t=1}^{T} \left( R_{t,F} - \sum_{i=1}^{N} R_{i,t} \times \frac{\left(S_{i,0} + \sum_{d=1}^{t-1} \Delta S_{i,d}\right) \times P_{i,t-1}}{\sum_{i=1}^{N} \left(S_{i,0} + \sum_{d=1}^{t-1} \Delta S_{i,d}\right) \times P_{i,t-1}} \right)^2 \tag{9}$$

subject to the constraints described in equations (5) and (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.

## 2.2 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 $k_i$:

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

$$k_i \in \{1, 2, 3, \dots, T\} \tag{10}$$

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

$$\min_{k_1 \dots k_N} \sum_{t=1}^{T} \left( R_{t,F} - \sum_{i=1}^{N} R_{i,t} \times \frac{S_{i,k_i,t-1} \times P_{i,t-1}}{\sum_{i=1}^{N} S_{i,k_i,t-1} \times P_{i,t-1}} \right)^2 \tag{11}$$

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.

### 2.2.1 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

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

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.

### 2.2.2 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:

$$flow_t = \frac{\sum_{i=1}^{N} S_{i,t} \times P_{i,t} - \left(\sum_{i=1}^{N} S_{i,t-1} \times P_{i,t-1}\right) \times (1 + R_{t,F})}{\sum_{i=1}^{N} S_{i,t-1} \times P_{i,t-1}}. \tag{12}$$

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:

$$\min_{k_1 \dots k_N} \sum_{t=1}^{T} \left[ \left( R_{t,F} - \sum_{i=1}^{N} R_{i,t} \times \frac{S_{i,k_i,t-1} \times P_{i,t-1}}{\sum_{i=1}^{N} S_{i,k_i,t-1} \times P_{i,t-1}} \right)^2 + 0.01 V_t \right], \tag{13}$$

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 $(k_1, \dots, k_N)$ according to the genetic algorithm.

## 2.3 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 (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

[4] The genetric 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).

between 1 and T. After repeating the procedure for 300 generations, I select the solution with the highest fitness.

# 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 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.

[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 et al. (2016) have used Ancerno outside of the mutual fund context.

## 3.1 Empirical Validation

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. 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 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:

$$Error_t = \sum_{i=1}^{N} \left( \frac{(S_{i,t,M} - S_{i,t,O}) \times P_{i,t}}{TNA_t} \right)^2, \tag{14}$$

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:

$$MSE = \frac{1}{T} \sum_{t=1}^{T} Error_t. \tag{15}$$

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 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.

### 3.1.1 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 (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:

$$ln(MSE_{holdings}) = \alpha + \beta \times ln(fitness_{model}) + \gamma' Controls + \epsilon. \tag{16}$$

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 2 presents the regression results of equation (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 (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 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%.

### 3.1.2 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

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

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 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 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.

# 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.

## 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.

[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.

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:

$$R_{t,F} = \frac{R_{t,F,reported}}{(1 - \%Cash)} \tag{17}$$

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 (13) with the right hand side of equation (17):

$$\min_{k_1 \ldots k_N, \%Cash} \sum_{t=1}^{T} \left[ \left( \frac{R_{t,F,reported}}{(1 - \%Cash)} - \sum_{i=1}^{N} R_{i,t} \times \frac{S_{i,k,t-1} \times P_{i,t-1}}{\sum_{i=1}^{N} S_{i,k,t-1} \times P_{i,t-1}} \right)^2 + 0.01 V_t \right], \tag{18}$$

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.

## 4.2 Liquidity Provision of Active Mutual Funds

In order to estimate the daily trades of US active mutual funds, I minimize equation (18) using the genetic algorithm. To reduce noise in the estimates, I restrict the sample to funds

[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.

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 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 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.

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

[10] Fitness is defined as previously: the objective function (equation 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.

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.

### 4.2.1 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 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 $Mom1\&5_{calc}$.

### 4.2.2 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:

$$R - R_f = \alpha + \beta \times Liquidity\ Provision + \gamma' Controls + \epsilon. \quad (19)$$

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 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.

### 4.2.3 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 $mom1\&5_{calc}$:

$$flow_t = \beta_0 + \beta_1 \times mom1\&5_{calc} + \gamma' Controls + \epsilon. \tag{20}$$

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 $\times$ style dummies. Overall these measures control for the fund flows that result from performance without an explicit recognition of liquidity provision.

Table 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.

# 5 Conclusions

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.

# References

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

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

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.

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, 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.

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.” *Unpublished working paper. University of Lugano and Swiss Finance Institute.*

Frazzini, Andrea and Owen A Lamont (2008), “Dumb money: Mutual fund flows and the cross-section of stock returns.” *Journal of financial economics*, 88, 299–322.

Ha, JinGi and Jianfeng Hu (2017), “How smart is institutional trading?” *Working Paper.*

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

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.”

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

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*.

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.

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.

Table 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 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 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 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 (×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 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 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 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 $\times$ 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 (%) |
|---|---|---|
| $mom1\&5_{calc}$ ($\times$ 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: 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.

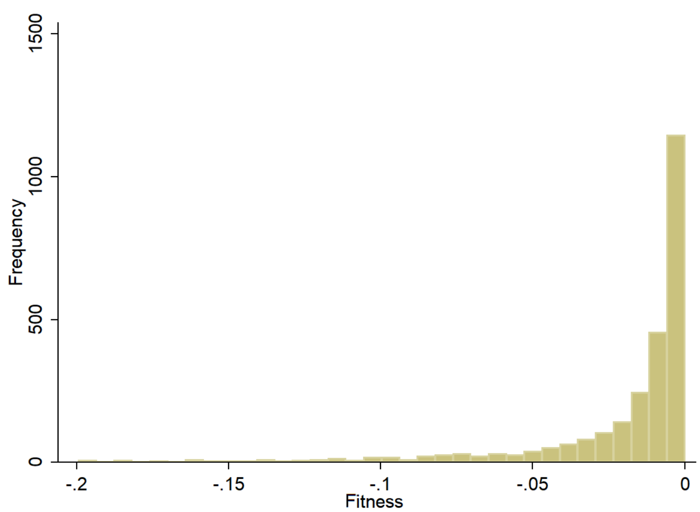

Figure 2: $Mom1\&5_{calc}$ Persistence

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.

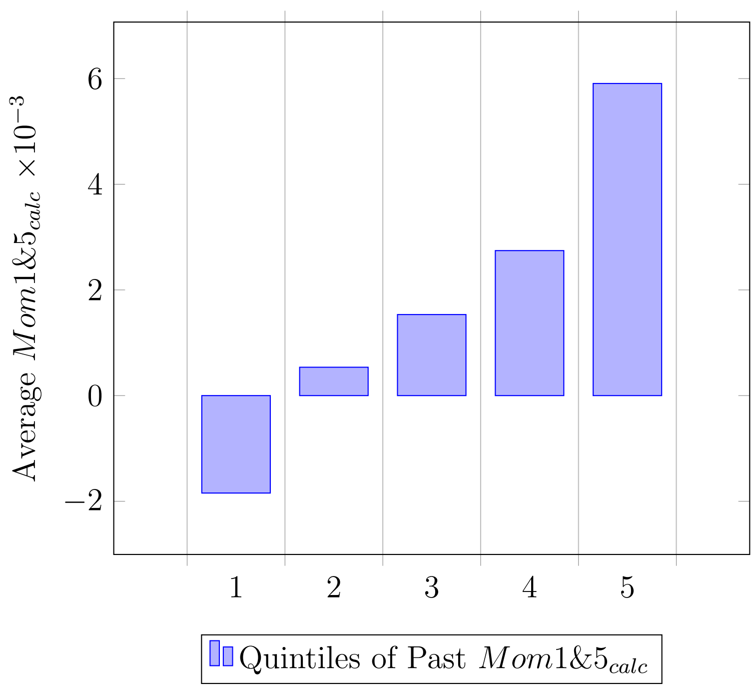

Figure 3: $Mom1\&5_{calc}$ and Future Returns

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.