The Daily Trading of Mutual Funds

Stephen P. Boyd*, Veronika K. Pool†, Noah Stoffman†, and Scott E. Yonker‡

Abstract

Mutual funds have been much studied in finance due to the critical role they play in financial markets. But this vast literature suffers from a major shortcoming: fund behavior is observed only irregularly. That is, while funds report returns daily, their complete holdings are typically reported only at a quarterly, or perhaps monthly, frequency. We aim to fix this by applying machine learning techniques to uncover the daily trading activity of funds.

U.S. mutual funds manage over 19 trillion dollars in assets and provide access to financial markets to over 100 million U.S. investors.[1] A vast literature exploring the behavior of funds and their impact on financial markets dates back to Jensen's (1968) seminal paper. But the daily trading behavior of mutual funds is unobserved.

Several papers have used clever empirical specifications to attempt to identify some of the implications of these hidden trades, including Kacperczyk, Sialm, and Zheng (2006), Sias, Starks, and Titman (2006), and Campbell, Ramadorai, and Schwartz (2009). In this project, we go further by attempting to use available information to infer the actual daily trading of mutual funds.

Our approach is as follows. We observe the complete holdings, H_t, of a fund at the beginning ($t = 0$) and end ($t = T$) of the period (either a month or quarter). We also observe daily fund returns and stock prices. The problem is to identify H_t for all $t \in \{1, \ldots, T-1\}$. Denote by $\theta_{i,t}$ a fund's trades in stock i on date t, so

$$H_{i,t} = H_{i,0} + \sum_{\tau=1}^{t} \theta_{i,\tau}.$$

Given intra-period trading, $\Theta \equiv \{\theta_{i,t}\}_{i=1,t=1}^{N,T-1}$, we can calculate the holdings on any date t, and, by combining with observed daily stock prices, an implied fund TNA.[2] Trades can therefore be estimated by solving

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \sum_{t=1}^{T-1} (V_t - V_t(\Theta))^2, \tag{1}$$

where V_t denotes TNA. There are an infinite number of solutions to this problem,[3] but an economically meaningful solution must be *sparse*, meaning that most of the elements of Θ are zero; funds are highly unlikely to trade most of their stocks on most days.

*School of Engineering and Inst. for Computational and Mathematical Engineering, Stanford University

†Kelley School of Business, Indiana University

‡Dyson School of Applied Economics and Management, Cornell University

[1] `https://www.ici.org/research/stats/trends/trends_02_19`

[2] While actual trading prices are not observable, we can use the daily volume-weighted average price as a proxy. This formulation also abstracts away from issues related to investor flows, which can be mitigated by using data from Trimtabs. Moreover, daily TNA is not typically observable but may be approximated using fund returns and the previous end-of-period reported TNA.

[3] $V_t - V_t(\hat{\Theta})$ can be written as $A\theta = b$ where A has dimension $(T-1) \times [N \times (T-1)]$.

A machine learning heuristic for finding such a sparse solution is "norm approximation" using regularization (Boyd and Vandenberghe, 2004, chapter 6). By adding an ℓ_1-norm penalty function to the minimization, we force the algorithm to identify a sparse solution:

$$\underset{\Theta}{\text{minimize}} \, ||V_t - V_t(\Theta)||_2^2 + \lambda ||\Theta||_1,$$

where $||\cdot||_k$ denotes the ℓ_k-norm. This particular regularization is related to Lasso regression, but other penalty functions, such as Elastic net could also be used. We have also investigated using a metaheuristic such as a genetic algorithm, although the benefit of using the regularization approach is that we are able to exploit the convexity of the problem to efficiently find the unique global minimum.

Additional economically-motivated constraints on trading can easily be added. For example, funds likely attempt to minimize transaction costs while trading. A standard model (Boyd et al., 2017) for the cost of trading x dollars of stock i is

$$TC_i(x) = a|x| + b\sigma_i \frac{|x|^{3/2}}{V_i^{1/2}} + cx,$$

where a is the half-spread, and σ_i and V_i are stock i's volatility and the total value of trading, respectively. Incorporating these costs into the the objective function would tend to deliver a solution that is not only sparse, due to the ℓ_i-norm component of $TC(x)$, but also one that identifies trading opportunities that may have been exploited by the fund. Other models, such as Kyle and Obizhaeva (2016), could also be incorporated.

We test our approach in two settings. First, we use simulated stock returns and funds to generate a series of periodically-reported fund holdings and daily fund returns and test our algorithm's ability to uncover the "true" daily holdings. Second, we are able to identify a small number of funds in the Ancerno data (Hu et al., 2013), where we can observe the complete trading of the fund. Again, we can compare our algorithm's ability to identify the actual trading days.

Our estimated trading results will be useful in any subsequent work on fund behavior, and we plan to make available data to researchers. While the data will be generally useful, it is also worth noting that they will be particularly useful in situations where stock-level aggregate fund trading is needed. That is, since our approach yields independent estimates for each fund's trading, aggregating the trading estimates across funds will yield more precise estimates of trading for a given stock.

Computing resources

We have already implemented parts of our algorithm in Python, using the `cvxpy` library, which uses the method of Disciplined Convex Programming (Grant and Boyd, 2008, 2014). We will use High-performance computing resources at Indiana University ("Big Red II").

References

Boyd, S., E. Busseti, S. Diamond, R. N. Kahn, K. Koh, P. Nystrup, J. Speth, et al. 2017. Multi-period trading via convex optimization. *Foundations and Trends in Optimization* 3:1–76.

Boyd, S., and L. Vandenberghe. 2004. *Convex optimization*. Cambridge University Press.

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

Grant, M., and S. Boyd. 2008. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, eds., *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, 95–110. Springer-Verlag Limited.

———. 2014. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`.

Hu, G., R. D. McLean, J. Pontiff, and Q. Wang. 2013. The year-end trading activities of institutional investors: Evidence from daily trades. *Review of Financial Studies* 27:1593–614.

Jensen, M. C. 1968. The performance of mutual funds in the period 1945–1964. *Journal of Finance* 23:389–416.

Kacperczyk, M., C. Sialm, and L. Zheng. 2006. Unobserved actions of mutual funds. *Review of Financial Studies* 21:2379–416.

Kyle, A. S., and A. A. Obizhaeva. 2016. Market microstructure invariance: Empirical hypotheses. *Econometrica* 84:1345–404.

Sias, R. W., L. T. Starks, and S. Titman. 2006. Changes in institutional ownership and stock returns: Assessment and methodology. *Journal of Business* 79:2869–910.