# Tomorrow Is Another Day: Stocks Overweighted by Active Mutual Funds Predict the Next-Day Market *

Shuaiyu Chen† Yixin Chen‡ Randolph B. Cohen§

First Draft: August 16, 2021
This Version: February 12, 2022

**Abstract**

We show that active mutual fund managers effectively incorporate information about future short-term market movements into security prices. Specifically, when high active-mutual-fund ownership stocks outperform, the market tends to do well the next day, and vice-versa. These effects are modest day by day but are quite large in the aggregate - trading the S&P 500 futures daily based on the strategy delivers an average annual return over 15% with a Sharpe ratio over 0.9. The same findings are also present in other major equity markets all around the world. Various additional tests further suggest that the novel short-term market return predictability results from active mutual fund managers' collective information advantage about future market movements, as opposed to informed fund flows or temporary price pressure.

*Keywords:* Market Predictability, Active Mutual Fund, Market Efficiency
*JEL Codes:* G10, G14, G17, G23

*We appreciate valuable discussions with Joshua Coval, Zhi Da, Robin Greenwood, Ron Kaniel, Alan Moreira, Robert Novy-Marx, Christian Opp, Christopher Polk, Charles Trzcinka (discussant), Tuomo Vuolteenaho, and Jerry Warner. We also thank seminar and conference participants at University of Rochester brown bag and Wabash Finance Conference for helpful comments.

†chen4144@purdue.edu, Krannert School of Management, Purdue University.
‡yixin.chen@simon.rochester.edu, Simon Business School, University of Rochester.
§rbcohen@hbs.edu, Harvard Business School, Harvard University.

# 1. Introduction

At the heart of the study of finance lies the question: how informationally efficient is the stock market? Important sub-issues include: how efficient is the price of the market as a whole, and how efficiently are individual stocks priced relative to one another? How efficient are prices at shorter frequencies such as a minute or a week, and how efficient are they at longer horizons such as a year or a decade? Moreover, what are the forces driving prices toward, and away from, fair value? What role do Wall Street's high-paid professionals play in the process of getting prices right?

This paper looks primarily at the efficiency of the pricing of the entire market, though we rely on, and have some findings concerning, pricing of different types of stocks as well. The frequency of our interest is daily. The paper's core finding is the following: actively-trading investment professionals such as mutual fund managers appear to process and exploit considerable information about expected prices over the next day. When these professionals perceive that the market, or the particular stocks they are interested in, will perform poorly in the future, they incorporate such information into their holdings, pushing down the prices of those stocks relative to those of firms with lower ownership from active institutions (we dub this measure "active ownership" or AO). When that occurs, the market has a marked tendency to perform poorly over the next day. Similarly, when high-AO firms perform better than low-AO firms, the market tends to do well over the subsequent trading day.

This empirical finding motivates us to build a market-timing strategy trading the S&P 500 futures on a daily basis. We find that the strategy delivers an average annualized return of more than 15% with a Sharpe ratio of 0.95. In addition, the strategy exhibits exceptionally appealing performance during major market downturns, such as the dot-com bubble bust, the 2008 Great Recession, and the recent Covid-19 episode.

Furthermore, we verify that our novel finding on the predictability of the short-term market return by the relative performance between high-AO and low-AO stocks is not limited to the US equity market; but rather, it is a robust phenomenon present in some of the most important equity markets around the world (e.g., Great Britain, China, Japan, etc).

Our key findings suggest that, first, the overall market may be less efficiently priced than previously suspected, and, second, active managers may play a crucial role in reducing those inefficiencies at the whole-market level.

We next explore the underlying mechanism driving this short-term market return predictability. We hypothesize that our findings are due to active managers as a group being traders informed about price movements over the next day or two. Although our tests show that the relative performance of high- and low-AO stocks predicts the entire market, it is not clear that any individual manager has market-timing ability; we can say only that the market can be predicted by *aggregating* the information in active-manager decisions in the whole cross section. While it is possible that managers have market-predicting insights that are successfully driving their trades, our findings could well be generated by managers who have tradable information only about individual stocks they are involved in, not about the market or economy as a whole, and indeed the latter is the basis on which we model the phenomenon. Also worth recognizing is the fact that a large body of empirical research has shown that the active mutual fund industry only generates very modest and insignificant pre-fee alpha in aggregate (see, Malkiel, 1995; Fama and French, 2010). The average pre-fee alpha in active-manager portfolios, which is close to zero in our sample, may substantially underestimate the information advantages of the active managers and the efficiency gains created by their active trading. That is because, if the informed trades mostly occur among active managers, then the profits generated by the informed funds would be offset by the losses of the uninformed counter-party funds.

To clarify the mechanism by which active managers can substantially improve market efficiency while generating little or no alpha for the active-management group, it helps to consider a fictitious example where all companies issue two types of shares: A-shares and B-shares. The A-shares are held by and traded among active mutual funds, whereas other investors own the B-shares. Suppose the active mutual fund managers are collectively more informed about next-day market prospects than the other investors, then the A-shares would incorporate such market-wide information earlier than the B-shares, and the relative performance between these two groups of stocks would predict the market return for one day. Notice that the active managers do not trade with the other investors, but they can still incorporate information into stock prices by trading among themselves. Moreover, no individual manager need trade all the A-shares; each might only trade a handful, and it could still be the case that the combined A-share trading of all active managers leads to strong predictions of the market. In this example, even though the active mutual funds are collectively more informed than the other investors, they may generate little or no alpha in aggregate; the profits generated by the informed fund managers equal the losses of the uninformed managers who act as the counter-parties, so trading profits for these managers are a zero-sum game. However, this is not necessarily a zero-sum game for welfare, as they could improve the information efficiency of the stock prices via their trading and, in turn, have a positive impact on the real economy (French, 2008).

This example is, of course, over-simplified and abstracts away from many realistic features of the market. Nevertheless, it serves as a useful framework that captures the essence of our hypothesis. In Section 2 we build a dynamic Grossman-Stiglitz-type model to illustrate how each individual manager can incorporate aggregate news into stock prices when trading individual stocks.

In subsequent tests, we entertain several alternative hypotheses, and we find that the empirical evidence mostly supports our channel. First, we confirm that our main findings are indeed driven by active-mutual-fund ownership as opposed to confounding firm- or stock-level characteristics related to liquidity or visibility (e.g., analyst and media coverage). As another competing explanation, it could be that active managers do net buying on certain days not because they foresee a rising market, but simply because they received investor inflows. Indeed that is sure to be one driver of active-manager trading. We test to see if it is the investor flows, rather than active-manager opinion, that drives market predictability. This appears not to be the case since passive funds have flows too, but the price movements in the stocks held by passive funds do not predict the next-day stock market. We also examine different kinds of professional investors. Consistent with our hypothesis, we find only holdings of active management companies (i.e., active mutual funds or hedge funds) exhibit predictive power; whereas the holdings of other institutional investors such as banks and insurance companies do not contribute to short-term market return predictability at all. In the same spirit, within the active mutual fund sector, we demonstrate that the predictability of our signal is mostly attributable to those funds with better performance, higher trading activity (Kacperczyk et al., 2008; Pástor et al., 2017), and more concentrated portfolio (Kacperczyk et al., 2005; Cremers and Petajisto, 2009). Going beyond stock-price predictions, we show that, consistent with our hypothesis but not predicted by many other potential explanations, the high-AO/low-AO performance gap also predicts the next-day market sentiment aggregated from news articles of individual firms.

To supplement our key observations regarding market return predictability, we also look inside the market to see if the relative performance of high- and low-AO stocks can be used to predict industry performance as well as that of the whole market; we find that this is indeed the case, in particular for the industries with high active-mutual-fund participation. Lastly, at the individual stock level, similar to Hameed et al. (2017), we find that it is the returns of the high-AO stocks that lead the returns of the low-AO stocks, and not vice versa; which is, again, consistent with our hypothesis. These various observations all point in the same direction and provide strong support for our hypothesis regarding the collective information advantage of the active mutual fund sector.

In sum, our paper identifies a large empirical anomaly in the pricing of the entire stock market, suggesting significant short-term price predictability. Moreover, we find that mutual funds and other active investors play a significant role in resolving market mispricings over time horizons of a day or two.

Our paper contributes to two strands of the finance literature. First, we document novel predictability of the short-term stock market return that is prevalent around the world. Voluminous research has documented return predictability in the cross section of stocks (see Lewellen (2014); McLean and Pontiff (2016); Hou et al. (2020) and literally thousands of others); Yet, much less is documented regarding the predictability of the entire stock market, especially in the short term. Of course there are a handful of such results; notable examples include: Lakonishok and Smidt (1988), Savor and Wilson (2013), Lucca and Moench (2015), Chen et al. (2020) showing abnormal market performance on certain pre-determined dates; Bollerslev et al. (2009), Ross (2015), Martin (2017) relating market return predictability to the pricing of derivative contracts; Campbell et al. (1993), Kelly and Pruitt (2013), Huang et al. (2015), Rapach et al. (2016), Jiang et al. (2019), Engelberg et al. (2019), Dong et al. (2021) predicting the market return with various market conditions. Our finding of the one-day market return predictability by AO-sorted stocks is a new addition to this literature. Second, our paper contributes to the large literature studying the skills of mutual funds, especially fund managers' ability to time the market (see, for example, Henriksson and Merton (1981); Bollen and Busse (2001); Jiang et al. (2007); Kacperczyk et al. (2014)). Our paper is most closely related to Bollen and Busse (2001), which also studies the market-timing abilities of mutual fund managers at daily frequency. Our paper differs from theirs in that they focus on fund performance due to managers' market-timing skills, whereas we highlight the market-wide information that is incorporated into securities prices by the funds. As demonstrated in our simple example above, these two effects are conceptually different and need not co-exist. In other words, we show that the whole active mutual fund industry incorporates considerable market-wide information into security prices, even though they generate an average pre-fee alpha close to zero. Also related to our paper, Hameed et al. (2017) explains the lead-lag relation between large and small stocks with institutions' slow trading behavior. Similar to their paper, we also discovered cross-predictability across stocks sorted by active-mutual-fund ownership. But our paper differs from theirs as we mostly focus on market return predictability instead of individual stocks, and we show that the novel predictability associated with active ownership is not attributed to confounding firm characteristics such as size or liquidity, which play a central role in their study.

The rest of the paper is organized as follows. Section 2 presents a dynamic asymmetric information model that motivates our empirical exercises; Section 3 introduces the data employed in our study; the empirical results are all contained in Section 4; Section 5 concludes.

## 2. Theoretical Framework

Motivated by Grossman and Stiglitz (1980) and Wang (1993), we consider a dynamic model with asymmetric information to illustrate how active managers with only stock-level information can collectively incorporate aggregate information into security prices. The model clarifies the theoretical framework that we employ to guide our empirical investigation and explain the findings.

### *2.1. Overview*

There are two types of investors in the model: a group of informed active managers trading individual stocks and *one* uninformed investor passively holding a diversified portfolio. The market features perfect segmentation: the group of active managers collectively trade and hold a set of high-AO stocks, whereas the uninformed investor holds the set of low-AO stocks.

Each informed active manager studies a particular stock. The signal received by the manager is about the total risk (including both the aggregate and idiosyncratic components) of a specific stock but *not* about the aggregate market directly. Each manager then takes advantage of the stock-level information that is available to her, and is only active at trading and incorporating her information into a particular stock. In the meantime, since the stock itself loads on the aggregate shock, the private signal received by each manager also contains a small aggregate component; via aggregation of all the high-AO stocks, the collective actions of all active managers incorporate aggregate information into the price of the high-AO portfolio, where the stock-specific shocks are diversified away.

The price of the high-AO portfolio is the source of aggregate information for the uninformed investor, who then incorporates such information into the low-AO stocks. However, due to the presence of noise demand, the price of the low-AO portfolio is not fully revealing and is thus less informative than the high-AO portfolio.

Finally, subtracting the low-AO portfolio's return from the high-AO portfolio's return removes common components that affect the contemporaneous market but have no predictive power for the future, so the relative performance between the two is a strong signal that predicts the next-period market.

## 2.2. Model Setup

We can consider a dynamic setting with infinite periods, i.e. $t = 0, 1, 2, \ldots$.

### 2.2.1. Securities

The market is segmented with two groups of risky securities. There is a group of $N$ stocks held by informed active managers (the high-AO stocks), as well as a group of $N$ stocks held by the uninformed investor (the low-AO stocks). Each stock pays out a stream of dividends:

$$\begin{aligned} & D_{i,t}^{Hi} = \mu + d_{a,t} + d_{i,t}^{Hi}, \\ \text{or} \quad & D_{i,t}^{Lo} = \mu + d_{a,t} + d_{i,t}^{Lo}, \end{aligned} \tag{1}$$

where $D^{Hi}_{i,t}(D^{Lo}_{i,t})$ denotes the dividend paid at date $t$ by stock $i$ in the high-AO (low-AO) group; $\mu > 0$ is the unconditional expected dividend payment; $d_{a,t+1} = \phi d_{a,t} + \epsilon_{a,t+1}$ with $\phi \in (0,1)$ and $\epsilon_{a,t+1} \sim \mathbb{N}\left(0, \sigma_a^2\right)$ is the aggregate component in the dividend process; and $d^{Hi}_{i,t+1} = \phi d^{Hi}_{i,t} + \epsilon^{Hi}_{i,t+1}$ $\left(d^{Lo}_{i,t+1} = \phi d^{Lo}_{i,t} + \epsilon^{Lo}_{i,t+1}\right)$ with $\epsilon^{Hi}_{i,t+1} \sim \mathbb{N}\left(0, \sigma_i^2\right)$ $\left(\epsilon^{Lo}_{i,t+1} \sim \mathbb{N}\left(0, \sigma_i^2\right)\right)$ is the stock-specific component. The random cash flow shocks $\left\{\epsilon_{a,t}\right\}_{t=0}^{\infty}, \left\{\epsilon^{Hi}_{i,t}\right\}_{t=0}^{\infty}, \left\{\epsilon^{Lo}_{i,t}\right\}_{t=0}^{\infty}$ are i.i.d. normal.

All risky securities are of unit supply. In addition to the risky securities, a risk-free asset is also available to all agents with a fixed interest rate $R > 1$.

### *2.2.2. Agents*

There are two types of long-lived investors in the market: a group of $N$ informed active managers, and *one* uninformed investor. All investors are endowed with the same initial wealth $W_0$ and have the same preference. Each investor maximizes over

$$\begin{aligned} J\left(W_t\right) &= \max_{\left\{C_s, \vec{X}_s\right\}} \sum_{s=t}^{\infty} \beta^{s-t} \mathbb{E}\left(U\left(C_s\right) | \mathcal{F}_t\right) \\ \text{s.t. } C_s + \vec{X}_s' \cdot \vec{P}_s &\leq W_s \text{ and } W_{s+1} = \vec{X}_s' \cdot \left(\vec{P}_s + \vec{D}_s\right) \end{aligned} \tag{2}$$

where

$$\begin{aligned} U\left(C\right) &= -\exp\left(-\alpha^I C\right), \\ \text{or} \quad U\left(C\right) &= -\exp\left(-\alpha^U C\right), \end{aligned} \tag{3}$$

is the per period utility; $W_t$ is the wealth available to the investor at time $t$; $C_t$ is the investor's consumption choice at $t$; $\vec{X}_t$ represents the number of stock shares in the investor's portfolio; $\vec{P}_t$ is the vector of stock prices; $\beta \in (0,1)$ is the subject discount factor; $\alpha^I(\alpha^U)$ is the absolute risk aversion of the informed (uninformed) investor; $\mathcal{F}_t$ represents the information set that is available to the investor at $t$; and $J(\cdot)$ is the value function.

Each informed active manager $i$ can only invest in the $i$th security of the high-AO group and the risk-free asset. And in each period, the manager receives a private signal of the security's next-period payoff:

$$s_{i,t} = \epsilon_{i,t+1}^{Hi} + \epsilon_{a,t+1}. \tag{4}$$

Note that the private signal fully reveals the next-period cash flow of security $i$; since the active manager can only trade one stock, given the signal, she no longer needs to infer information from stock prices.[1]

The uninformed investor is able to invest in all securities of the low-AO group and the risk-free asset, but she cannot invest in the securities in the high-AO group.

*2.2.3. Noise Demand*

There is noise demand for each security in both groups,

$$\begin{aligned} u_{i,t}^{Hi} &= u_t^{Hi} + \eta_{i,t}^{Hi}, \\ u_{i,t}^{Lo} &= u_t^{Lo} + \eta_{i,t}^{Lo}, \end{aligned} \tag{5}$$

where $u_t^{Hi} \sim \mathbb{N}\left(0, \sigma_u^2\right)$ $\left(u_t^{Lo} \sim \mathbb{N}\left(0, \sigma_u^2\right)\right)$ is the aggregate noise demand of the high-AO (Low-AO) group; and $\eta_{i,t}^{Hi} \sim \mathbb{N}\left(0, \sigma_\eta^2\right)$ $\left(\eta_{i,t}^{Lo} \sim \mathbb{N}\left(0, \sigma_\eta^2\right)\right)$ is the stock-specific noise demand. The random noise demand shocks $\left\{u_t^{Hi}\right\}_{t=0}^{\infty}, \left\{u_t^{Lo}\right\}_{t=0}^{\infty}, \left\{\eta_{i,t}^{Hi}\right\}_{t=0}^{\infty}, \left\{\eta_{i,t}^{Lo}\right\}_{t=0}^{\infty}$ are all i.i.d normal.

[1]The full-information-revelation assumption simplifies the solution of the model. If the private signal is not fully revealing, then the informed active manager needs to learn from both her private signal and the aggregate market price to extract additional information about $\epsilon_{a,t+1}$. The model will still be tractable, but it will feature tedious algebra while delivering the same insight.

## 2.3. Equilibrium

**Proposition 1.** *The aforementioned economy features a linear equilibrium, where the price of a high-AO stock is*

$$P_{i,t}^{Hi} = \frac{1}{R-1}\left(\mu - A^{Hi}\right) + \frac{\phi}{R-\phi}\left(d_{a,t} + d_{i,t}^{Hi}\right) + B^{Hi}\left(\epsilon_{a,t+1} + \epsilon_{i,t+1}^{Hi}\right) + C^{Hi}u_{i,t}^{Hi}; \tag{6}$$

*and the price of a low-AO stock is*

$$\begin{aligned} P_{i,t}^{Lo} = & \frac{1}{R-1}\left(\mu - A^{Lo}\right) + \frac{\phi}{R-\phi}\left(d_{a,t} + d_{i,t}^{Lo}\right) + B^{Lo}\left(\frac{\sum_i d_{i,t}^{Lo}}{N} - d_{i,t}^{Lo}\right) \\ & + C^{Lo}\left(B^{Hi}\left(\frac{\sum_i \epsilon_{i,t+1}^{Hi}}{N} + \epsilon_{a,t+1}\right) + C^{Hi}\frac{\sum_i u_{i,t}^{Hi}}{N}\right) + D^{Lo}u_{i,t}^{Lo}; \end{aligned} \tag{7}$$

*for some positive constants: $A^{Hi}, B^{Hi}, C^{Hi}, A^{Lo}, B^{Lo}, C^{Lo}$, and $D^{Lo}$.*

*Proof.* See Appendix A. □

**Proposition 2.** *The average price of the high-AO stocks is more informed about the next-period aggregate shock $\epsilon_{a,t+1}$ than the average price of the low-AO stocks, i.e.,*

$$Var\left(\epsilon_{a,t+1}|P_t^{Hi}\right) < Var\left(\epsilon_{a,t+1}|P_t^{Lo}\right), \tag{8}$$

*where $P_t^{Hi} \equiv \frac{1}{N}\sum_i P_{i,t}^{Hi}$ and $P_t^{Lo} \equiv \frac{1}{N}\sum_i P_{i,t}^{Lo}$ are the average prices of the high-AO and low-AO stocks.*

*Proof.* See Appendix A. □

**Proposition 3.** *Define signal $s_t$ as the return difference between the high-AO and low-AO stocks, i.e.,*

$$s_t \equiv R_t^{Hi} - R_t^{Lo}, \tag{9}$$

*where $R_t^{Hi} \equiv P_t^{Hi} + D_t^{Hi} - P_{t-1}^{Hi}$ and $R_t^{Lo} \equiv P_t^{Lo} + D_t^{Lo} - P_{t-1}^{Lo}$ are the average (dollar) returns of the high-AO and low-AO stocks.*

*Under the parameter specification where $Var\left(R_t^{Hi}\right) = Var\left(R_t^{Lo}\right)$ and $N$ is large,*

$$s_t \approx \left(1 - C^{Lo}\right) B^{Hi} \left(\epsilon_{a,t+1} - \epsilon_{a,t}\right) + \left(1 - C^{Lo}\right) C^{Hi} \left(u_t^{Hi} - u_{t-1}^{Hi}\right) + D^{Lo} \left(u_t^{Lo} - u_{t-1}^{Lo}\right) \tag{10}$$

*for some $C^{Lo} \in (0, 1)$.*[2]

*Therefore, $s_t$ is predictive of the next-period aggregate shock $\epsilon_{a,t+1}$, and thus the next-period market return.*

*Proof.* See Appendix A. □

# 3. Data

The data in this paper are from several sources. The US and global equity data come from CRSP and Compustat Global, respectively. We include firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ that have a CRSP share code of 10 or 11. For the global data, we include common stocks listed in the 10 largest equity markets: the United States, Japan, China, Great Britain, Hong Kong, France, German, Canada, India, and Switzerland. Stock share prices and returns are converted into US dollars using the exchange rates from Compustat.

To track the performance of the aggregate equity market, we collect the daily prices of the most liquid and traded futures for the following 10 equity market indices: S&P 500 (United States), TOPIX (Japan), CSI 300 (China), FTSE 100 (Great Britain), HSI (Hong Kong), CAC 40 (France), DAX (German), TSX (Canada), NIFTY 50 (India), and SMI (Switzerland). These market index futures are commonly used in the asset pricing literature to study the behavior of market returns (see Moskowitz et al., 2012; Koijen et al., 2018).

The data on US equity mutual funds are from the CRSP Survivor-Bias-Free US Mutual Fund Database and Thomson Reuters S12 Mutual Fund Holdings Database. We include actively managed mutual funds, index funds, and exchange-traded funds

[2]Empirically, the daily volatility of the high-AO portfolio is close to the low-AO portfolio (1.31% VS 1.15%).

(ETFs). The active funds are identified based on the screening procedure used in Kacperczyk et al. (2008) and Cremers and Pareek (2016).[3] To identify index funds and ETFs, we first rely on the fund type indicator in CRSP, then screen by fund names following the procedure proposed by Appel et al. (2016).[4] To mitigate the incubation bias highlighted by Evans (2010), we include a fund in our sample after its inception date and when its total net assets first pass $5 million in 2006 dollars (Fama and French, 2010). Zhu (2020) documented that, from 2010 to 2015, 58% of newly founded US equity mutual fund share classes in CRSP cannot be matched with the Thomson Reuters database. To deal with this data issue, we retrieve mutual fund holdings from Thomson Reuters before June 2010 and from CRSP afterwards.[5] For funds with multiple share classes, we aggregate all share classes at the portfolio level. The final sample of US equity mutual funds includes 5,810 actively managed funds, 688 index funds, and 793 ETFs.

We acquire 13F institutional holdings from both the Thomson Reuters S34 Holdings Database and the holdings data provided by Wharton Research Data Services' (WRDS) SEC Analytics. Ben-David et al. (2021) pointed out several data issues in the Thomson Reuters Database and assessed the potential biases caused by these issues. Following their suggestion, we use the Thomson Reuters data before June 2013 and the SEC 13F fillings data parsed by WRDS SEC Analytics afterwards.[6] We identify equity holdings of hedge funds based on the institution classification

[3] We evaluate the Lipper Prospectus objective code, the Strategic Wensight objective code, and the Weisenberger objective code to indicate that the fund is pursuing an active US equity strategy that does not focus on one or more particular industries or sectors. We require the Lipper Prospectus objective code to be EI, EIEI, ELCC, G, GI, LCCE, LCGE, LCVE, LSE, MC, MCCE, MCGE, MCVE, MLCE, MLGE, MLVE, MR, S, SCCE, SCGE, SCVE, SESE, SG, or missing; the Strategic Insight objective code to be AGG, GMC, GRI, GRO, ING, SCG, or missing; the Weisenberger objective code to be GCI, IEQ, IFL, LTG, MCG, SCG, G, G-I, G-I-S, G-S, G-S-I, GS, I, I-G, I-G-S, I-S, I-S–G, S, S-G-I, S-I, S-I-G, or missing; and the CDA/Spectrum code to be 2, 3, 4, or missing.

[4] We use the following strings in fund names to identify index funds: index, idx, indx, ind_ (where_ indicates a space), Russell, S & P, S and P, S&P, SandP, SP, DOW, Dow, DJ, MSCI, Bloomberg, KBW, NASDAQ, NYSE, STOXX, FTSE, Wilshire, Morningstar, 100, 400, 500, 600, 900, 1000, 1500, 2000, and 5000.

[5] Zhu (2020) showed that funds that are missing from the Thomson Reuters database tend to be smaller, have higher turnover, receive more fund flows, and have higher Carhart four-factor alphas.

[6] The data issues in the last few updates of the Thomson Reuters Database were discussed by the WRDS research team (https://wrds-www.wharton.upenn.edu/documents/752/Research_Note_-Thomson_S34_Data_Issues_mldAsdi.pdf).

in the FactSet Global Ownership Database (Ben-David et al., 2012).

Our data on global mutual fund holdings are from the FactSet Global Ownership Database, which has been widely used in recent studies on international mutual funds (see Cremers et al., 2016; Schumacher, 2018). The FactSet database covers various types of financial institutions (mutual funds, pension funds, investment advisors, etc.). We focus on the open-end mutual funds (OEF) and their equity holdings in the international study.

Finally, for the intraday analyses in our paper, we acquire the intraday transactions data from the NYSE Trade and Quote (TAQ) database. We construct analyst coverage and media coverage using the data from IEBS and RavenPack, respectively. Daily mutual fund flows are constructed using the data on daily total net assets from Morningstar Direct, available since August 2008. For most of the exercises on the US equity market, the sample period is from January 1983 to December 2020 when the S&P 500 index futures data are available; the sample starts from January 1987 (January 2001) when analyst (media) coverage is used. For the international analysis, the sample covers 2001 through 2020, during which the global mutual fund holdings are available in the FactSet Global Ownership Database.

# 4. Empirical Results

## *4.1. Main Findings*

### *4.1.1. Signal Construction*

Our key empirical finding is that the relative performance between stocks with high active-mutual-fund ownership and those with low active-mutual-fund ownership is predictive of the next-day market return. To extract this predictive signal, we take the following steps:

1. We exclude micro-cap stocks with market capitalization below the 20*th* percentile of NYSE stocks, following Fama and French (2008).

2. Among the remaining all-but-micro-cap stocks, we sort them into five groups based on active-mutual-fund ownership at the beginning of each quarter.[7]
3. We take the difference between the daily equal-weighted average returns of the high-AO stocks (group 5) and those of the low-AO stocks (group 1) as the *signal* to predict the market return over the next trading day.[8]

Our empirical procedure is motivated by several considerations. We exclude the micro-cap stocks because tiny stocks tend to exhibit erratic share price behavior (Fama and French, 2008; Hou et al., 2020), so we attempt to remove their influence on our signal. We take the equal-weighted averages of stock returns because the distribution of individual stock market capitalization is highly right-skewed in the cross section, meaning a market-cap weighting scheme would over-emphasize the firm-specific information of those mega-cap stocks. Notice that we do not trade the equal-weighted portfolios but only use them to extract a signal, so whether these equal-weighted portfolios are tradable is irrelevant to our study.[9] Lastly, motivated by our model, we take the difference in the average returns between the high-AO and low-AO groups to remove common market-wide components in stock returns that are not predictive of the subsequent market return, but could still contaminate our signal.[10]

#### *4.1.2. Predicting the Next-Day S&P 500 Futures Return*

To ensure that our finding of the daily market return predictability is fully tradable and not driven by potential market microstructure or liquidity issues, we adopt the S&P 500 futures as our main proxy of the US stock market. This empirical choice raises questions about whether our results are robust and how they would change if alternative financial instruments of the S&P 500 index were used. Table

[7]Lagging the holdings by two months has little effect on our results. See Table B3 and Table B4.

[8]Our empirical findings are not sensitive to the number of groups used in signal construction. These results are not tabulated but are available upon request.

[9]Rapach et al. (2016) implemented a similar equal-weighting scheme to predict the value-weighted market with a signal extracted from the short-interest value in a cross section of stocks.

[10]See Greenwood and Hanson (2012), Greenwood and Hanson (2013), Dong et al. (2021), etc. for similar procedures applied in predictive regressions.

B1 in Appendix B shows that our results are virtually unchanged when we consider alternative market proxies such as the E-mini S&P 500 futures, the S&P 500 ETF, the S&P 500 spot market, and the CRSP value-weighted market.[11]

**[Insert Table 1 near here]**

Table 1 presents our key findings. The table shows that the relative performance between the high-AO and low-AO stocks has significant power in predicting the next-day S&P 500 futures return. Column (1) shows that, when we regress the S&P 500 futures return on the one-day lagged signal, we obtain an OLS coefficient of 0.12, with a Newey and West (1987) $t$-statistic of 3.77, and the predictive regression has an R-squared of 0.20%.[12] Column (2) shows that controlling for the lagged market return further strengthens the result, where the coefficient on the signal rises substantially to 0.21 (with a $t$-statistic of 4.83) and the daily R-squared reaches an impressive level of 0.96%.[13] The remaining columns show that the predictive power of the AO signal is robust to monotonic transformations. In addition to the signal itself, both the positive and negative components, as well as the sign of the signal, strongly predict the next-day market return.

Table B2 in Appendix B further shows that our novel finding on daily market return predictability is not explained by the various market return predictors that have been proposed in the literature.[14]

[11]We consider the S&P 500 futures instead of the E-mini S&P 500 futures for our main analyses because the latter was launched in September 1997 and thus has a short sample period.

[12]To calculate Newey and West (1987) $t$-statistics, we choose the number of lags based on the rule-of-thumb $0.75T^{1/3}$, where $T$ is the number of trading days in the sample period.

[13]Previous studies documented a strong auto-correlation in market index returns. For example, Baltussen et al. (2019) found that return serial dependence in the S&P 500 futures was positive before the 1990s but switched to negative in the 2000s.

[14]We consider a long list of existing market return predictors, including the aggregate turnover (TO) (Campbell et al., 1993), Variance Risk Premium (VRP) (Bollerslev et al., 2009), Dividend Yield (DP), Earnings Yield (EP), Book-to-Market (BM), Inflation (INFL), Term Spread (TMS), Default Yield Spread (DFY), and Net Equity Expansion (NTIS) Welch and Goyal (2008).

*4.1.3. International Evidence*

To test the robustness of our findings, we apply our analysis to the 10 largest equity markets in the world: the United States, Japan, China, Great Britain, Hong Kong, France, German, Canada, India, and Switzerland. For each of these markets, we construct the signal in a similar way as described in Section 4.1.1, then study its predictability for the most traded and liquid market index futures (Moskowitz et al., 2012; Koijen et al., 2018).[15] Our sample is from 2001 to 2020, the period for which the FactSet Global Ownership Data are available.

**[Insert Table 2 near here]**

Similar to our main exercise with the US market, we predict the next-day futures return with the same signal, extracted from portfolios sorted by mutual fund ownership, for each of these equity markets. Table 2 shows that our novel daily market return predictability is significantly present in seven of the 10 largest equity markets, including the US, China, Japan, Great Britain, Canada, France and Switzerland.

Therefore, our key finding is not isolated to the US but is prevalent around the world. For the rest of the paper, however, we will focus on the US market due to data availability regarding stock characteristics, alternative financial institutions, and intraday trades and quotes.

*4.1.4. Economic Significance*

This strong daily S&P 500 futures return predictability is a striking result because the S&P 500 futures contract is very liquid and can be easily traded, both on the long and short. To illustrate the economic significance of our findings and evaluate the consistency of these effects, we construct a simple market-timing trading

[15] For each equity market, we first exclude micro-cap stocks with market capitalization below the $20th$ percentile of all stocks in the market, then sort the remaining stocks into two groups by active-mutual-fund ownership. The signal is computed as the difference between the daily equal-weighted average returns of the high-AO and low-AO stocks.

strategy based on our signal, then evaluate its performance using different factor models.

Following Campbell and Thompson (2008) and Gao et al. (2018), we construct the optimal portfolio for a mean-variance investor with a risk aversion coefficient of 5 using our active ownership signal. Specifically, our strategy adjusts the weight on the S&P 500 futures using the following formula:

$$w_t = \frac{\widehat{E}_t(r^e_{m,t+1})}{5 \times \widehat{Var}_t(r^e_{m,t+1})},$$

where $\widehat{E}_t(r^e_{m,t+1})$ is the out-of-sample expected return of the S&P 500 futures estimated with the data from July 1982 to the date of portfolio formation, and $\widehat{Var}_t(r^e_{m,t+1})$ is the out-of-sample variance estimated with a rolling window of 252 trading days. The weight is bounded between -0.5 and 1.5. The first seven years of the data are treated as a training sample, and the out-of-sample strategy starts from January 1990.

**[Insert Table 3 near here]**

**[Insert Figure 1 near here]**

Table 3 shows that our simple market-timing strategy delivers outstanding performance. Panel A of the table documents how, over the last 30 years, the strategy would realize a premium of 15% per year, with an impressive Sharpe ratio of 0.95. The utility gain to the mean-variance investor is equivalent to a 549 bps annual management fee to gain access to the strategy, with the alternative being predicting the market return by its historical mean. Panel B of the table shows that the large profitability of the strategy is not explained by its exposure to popular risk factors, and the information ratio of the strategy ranges from 0.78 to 0.85, depending on the benchmark. Figure 1 further shows that the attractive performance of our market-timing strategy is consistent throughout the sample and is not vulnerable to severe economic downturns or financial crises.

Therefore, the strong predictability of the S&P 500 futures return by our AO signal implies large trading profits and significant utility gains to investors in the stock market.

## *4.2. Evidence Inconsistent with Competing Explanations*

According to the dynamic asymmetric information model presented in Section 4.1, we conjecture that our strong next-day market return predictability is derived from an information channel, in which active managers incorporate aggregate information into security prices by picking and trading individual stocks. We first consider several alternative hypotheses in this subsection and show evidence that our finding is inconsistent with these channels. We then present the additional findings that can lend support to our preferred informational channel in the next subsection.

### *4.2.1. Market Return Predictability at Different Horizons*

In contrast to our conjectured asymmetric information mechanism, one competing explanation for the market return predictability is that it is a manifestation of a short-term price pressure which would quickly dissipate. It is conceivable that the good performance of the stocks favored by mutual funds would attract more fund flows, so that the managers would be forced to buy more stocks and push up the market price.

**[Insert Table 4 near here]**

To investigate this potential explanation, we study the horizon of our AO signal's predictability, both over the subsequent five days and within the next trading day. Panel A of Table 4 shows that the predictability of our signal lasts for one trading day with no subsequent reversal. We next decompose the signal and the market return into their intraday and overnight components[16]. Panel B of the table shows

[16]Following Bogousslavsky (2021), we take the price at 9:45am as the open price to mitigate po-

that our signal's predictability mainly stems from its intraday component on both the intraday and overnight components of the next-day market return.

The lack of subsequent reversals in stock price is inconsistent with the price pressure channel. In addition, the finding that the predictability of the signal only yields from its intraday component is consistent with the information channel, as this is the time window when mutual funds actively trade. It is not obvious how the price pressure channel would generate such a pattern, as mutual fund managers would still receive positive flows at the open and keep buying during the day if their portfolios realized good performances overnight.

### *4.2.2. Predictability by Alternative Firm Characteristics*

As discussed, the explanation of the market return predictability we find most consistent with the data is that active mutual fund managers are collectively informed, so that the prices of the stocks with high active-mutual-fund ownership adjust faster and predict the market. On the other hand, one competing hypothesis is that active-mutual-fund ownership is correlated with other types of firm characteristics such as liquidity or visibility, and that the stocks with these alternative confounding firm characteristics produce the signal that predicts the market.[17]

**[Insert Table 5 near here]**

Indeed, Table 5 shows that the active ownership measure has interesting relations with several firm characteristics. Consistent with the extensive literature on mutual fund portfolio preferences, stocks with high active ownership tend to be more liquid and have higher analyst coverage. The relation between market cap and active ownership exhibits an interesting inverted U-shape: stocks with extremely low and extremely high active ownership are smaller than the stocks in

tential microstructure issues; for the individual stock prices which we use to generate the signal, the open price is defined as the midquote at 9:45am.

[17] For the empirical relations between active mutual fund ownership and firm characteristics, see Falkenstein (1996), Bennett et al. (2003), Massa et al. (2004), Cao et al. (2013), Solomon et al. (2014), Fang et al. (2014), etc.

the middle. This is an intuitive finding because while mutual funds are reluctant to hold illiquid stocks which tend to be small, allocation to small stocks tends to result in high ownership.

**[Insert Table 6 near here]**

To rule out the possibility that our signal's predictability stems from these confounding firm characteristics, Table 6 presents the regressions using alternative signals constructed following the same procedure, but with the active ownership measure replaced by alternative firm characteristics. The table shows that only the signal generated by active-mutual-fund ownership demonstrates significant predictive power for the market. Therefore, we find that the predictability of our signal is not due to the active ownership measure's correlation with other confounding firm characteristics.

#### *4.2.3. Predictability by Alternative Financial Institutions*

In our preferred explanation, the daily market return predictability via the AO signal is a strong testament to the investment skills of active mutual fund managers. On the other hand, interesting questions arise regarding whether such skills are unique to active mutual funds and also whether they should be attributed to fund investors rather than fund managers. To investigate these questions, we conduct additional tests by applying our analysis to financial institutions with different business objectives and investment styles.

**[Insert Table 7 near here]**

In Table 7, we explore the differences across the seven major types of institutional investors: active mutual fund, passive fund (index fund and ETF), investment advisor, pension fund, bank, insurance company, and hedge fund. The table reproduces our main predictive regression with the signals constructed from stock ownership by alternative financial institutions. As expected, only the signals produced by active mutual funds, investment advisors (which include asset management companies), and hedge funds show predictive power for the market return.[18] Intuitively, ownership by pension funds, banks, and insurance companies does not help to predict the market return at daily frequency as these institutions do not take market-timing bets on a daily basis.

Importantly, the signal associated with ownership by passive funds and ETFs does not predict the market return either. Such a finding suggests that the predictive power of our signal derives from the market-timing skills of the active mutual fund managers instead of the flows from mutual fund investors, because the passive funds have flows too. We investigate the return predictability of fund flows directly in the next subsection.

It is also striking that there is a substantial difference in the market return predictive power between stocks with active ownership and passive ownership, even though active mutual funds and passive mutual funds on average generate similar pre-fee performance (Malkiel, 1995; Fama and French, 2010). Intuition would suggest that if active mutual funds' holdings lead other stocks, and the market as a whole, this will imply significant pre-fee performance for active funds relative to passive. But this need not be the case; if markets react quickly to the trades of active managers, they can be the channel by which information makes its way into the market while active managers receive only very modest performance benefits for their service (Grossman and Stiglitz, 1980; Kyle, 1985).

#### *4.2.4. Lack of Predictability by Aggregate Fund Flows*

To further rule out the possibility that the predictability of our signal derives from the market-timing skills of fund investors rather than fund managers, we follow Edelen and Warner (2001) to study the predictability between aggregate fund

[18]The table shows that the signal constructed with hedge fund holdings also significantly predicts the next-day market, but with a somewhat weaker magnitude than the signal derived from active-mutual-fund ownership. This weaker effect might be due to the fact that information about hedge fund holdings is less complete and accurate than mutual fund holdings, as hedge funds don't need to disclose their short positions and smaller hedge funds, i.e., those with AUM less than $100M, are not required to report their positions at all.

flows and the next-day market return.

**[Insert Table 8 near here]**

Using a proprietary dataset, Edelen and Warner (2001) showed that the aggregate daily mutual fund flow is concurrently correlated with market return but does not predict future return. In Table 8, we replicate Edelen and Warner's (2001) exercise in a more recent sample, using daily mutual fund flows derived from the Morningstar Direct data.[19] Specifically, Column (1) aggregates flows for all mutual funds in the cross section, Column (2) aggregates flows for all mutual funds that take the S&P 500 as benchmark, Column (3) aggregates flows for all active mutual funds, and Column (4) aggregates flows for all passive mutual funds. The table shows that, consistent with Edelen and Warner (2001), the aggregate flows received by funds in various universes do not forecast future market return.

Therefore, such a finding is inconsistent with the "informed fund investors" hypothesis, according to which aggregate fund flow should also predict future market return.

## *4.3. Evidence Consistent with the Information Channel*

### *4.3.1. Predictability of the Aggregate Stock News Sentiment*

In our model, we highlight the mechanism by which active mutual fund managers are able to incorporate aggregate information into security prices by trading individual stocks. To directly support our mechanism, we investigate our AO signal's ability to predict the next-day aggregated stock-level news sentiment.

**[Insert Table 9 near here]**

Specifically, for each day, we aggregate the stock-level news sentiment in the cross-section by taking the market-cap-weighted average of the Ravenpack Com-

[19] Morningstar Direct has provided daily total net asset data for a cross section of mutual funds since 2008. By merging Morningstar Direct with the CRSP mutual fund dataset, which provides daily mutual fund returns, we are able to infer daily mutual fund flows.

posite Sentiment Scores of news articles across all firms (dubbed the "aggregate news sentiment" or ANS).[20] We then show in Panel (A) of Table 9 that, consistent with our model, our AO signal is indeed able to predict the next-day aggregate stock news sentiment. Moreover, Panel (B) of the table further shows that the predicted aggregate stock news sentiment is significantly positively correlated with the next-day market return. These findings further support our information channel by showing that at least part of the AO signal's predictive power for the market return is derived from its ability to predict the next-day observable stock-level news, which is contemporaneously correlated with the next-day stock returns.[21]

### 4.3.2. *Predictability by Different Types of Active Mutual Funds*

To further support the information effects illustrated by our model, we explore the heterogeneity within the active mutual fund sector and show how the information channel successfully predicts the segments of the mutual fund industry that mostly contribute to the predictability of our signal.

**[Insert Table 10 near here]**

One implication of our model of active fund manager acumen is that the effect should be stronger if we isolate the managers who show the most evidence of investment talent. To test this conjecture, we partition the active mutual funds into two groups each quarter, based on the information ratio relative to the benchmark Carhart (1997) in the prior 24-month rolling window. Out-of-sample signals are produced by the ownership of these two groups of active funds separately. Column (1) in Table 10 shows that, consistent with our conjecture, the signal constructed from the funds with a high historical information ratio demonstrates much stronger predictive power than the signal produced from those with a low historical information ratio. The predictive power of the high-information-ratio signal is comparable

[20] We include all types of news articles except those in the "stock-prices" category.

[21] We aggregate the Ravenpack Composite Sentiment Score of all news articles except those belonging to the "stock-prices" topic group. Therefore, the news articles that we use are business-related and do not include reports directly about stock returns.

to the full version of the signal when all active mutual funds are included.

In the same spirit, since the predictability of our signal works on a daily frequency, we should also expect its effectiveness to come mostly from the funds that make high-frequency bets on the market. Previous studies have documented that high-turnover funds tend to outperform low-turnover funds because of their superior ability to exploit time-varying investment opportunities (Pástor et al., 2017). Motivated by this conjecture, we partition the active mutual funds into two groups by their prior-year turnover ratio, then construct separate signals based on high-turnover AO and low-turnover AO. Column (2) in Table 10 verifies this conjecture by showing that the predictability of the signal indeed derives from the high-turnover funds within the active mutual fund sector. Similarly, Column (3) shows that the funds with more profitable high-frequency trades (measured by return gap) contribute more to the predictability of the signal.[22]

In addition, motivated by Kacperczyk et al. (2005), Column (4) of the table compares the signals derived from funds with high or low industry concentration; Column (5) partitions funds by the active share metric proposed by Cremers and Petajisto (2009). These two columns further show that, consistent with our intuition, including funds that are more active (measured by either industry concentration or active share) generates a stronger predictive signal.

#### *4.3.3. Daily Industry Return Predictability*

Kacperczyk et al. (2005) documented that active mutual funds may possess private information about certain industries and consequently tilt their portfolio weights towards these industries. So to supplement our main empirical findings regarding market return predictability, we also extend the same logic and study the predictability of industry returns. To deviate from the market return predictability

[22] Kacperczyk et al. (2008) proposed the return gap measure to capture a mutual fund's ability to generate profits from unobservable within-quarter trades. The return gap of a fund in a specific month is defined as the difference between the actual performance of the fund and the counterfactual performance that the fund would have earned if it statically held the portfolio formed at the end of the previous quarter.

exercise, we measure the industry-specific returns by taking the difference between the daily value-weight industry returns and the market return. By the same token, we also produce industry-specific signals by only including stocks within a given industry when constructing the signals.

**[Insert Table 11 near here]**

Table 11 presents the predictive regressions with the industry-specific returns and signals. Motivated by Kacperczyk et al. (2005), we present the results of the table with industries ranked by their level of active mutual fund ownership, as we expect the effect to be stronger in industries with higher mutual fund ownership. Consistent with our intuition, we find strong industry-specific return predictability within industries featuring high active mutual fund ownership, such as finance or business services, but no predictability was found in industries with less mutual fund participation, such as telecom or utilities. The cross-industry findings complement our main results and suggest that the active mutual fund industry is collectively informed about systematic risks, both at the market level and at the industry level.

#### *4.3.4. Lead-Lag Relation by Active Mutual Fund Ownership*

We next extend our analysis to the full cross-section of stocks and study the lead-lag relations among their returns, in the same spirit as Lo and MacKinlay (1990) and Hameed et al. (2017). If the prices of high-AO stocks are indeed more efficient than other stocks, then we should expect cross-predictability in the returns of high-AO stocks to low-AO stocks, but not vice versa. To ensure such a lead-lag relation is tradable, we predict the stock returns from 9:45am to market close using the close-to-close returns on the previous trading day.[23]

**[Insert Table 12 near here]**

[23] Following Bogousslavsky (2021), we skip the first 15 minutes after market open to avoid potential microstructure/liquidity issues.

Table 12 confirms this conjecture and shows strong return cross-predictability among the stocks sorted by active ownership. The daily returns of the high-AO stocks positively predict the the low-AO stock returns. In contrast, the low-AO stock returns do not positively predict the high-AO stock returns. Instead, the coefficients of the low-AO stocks in the regression have slightly negative values because subtracting their returns from the high-AO stocks helps to remove the unpredictable market-wide common component and distill the signal that has the strongest predictive power.

The high-AO portfolio also demonstrates strong momentum at a daily frequency. Such a finding is consistent Lo and MacKinlay (1990) with the channel as that the lead-lag relation within the high-AO group can generate a momentum effect for the group as a whole.

## *4.4. Summary of the Mechanism*

Our empirical exercises reveal that the prices of the stocks with high active-mutual-fund ownership adjust faster than the rest of the market, and thus contain a signal that is predictive of the next-day's market or industry return. We summarize below the potential mechanisms underlying our empirical findings.

We speculate that active mutual fund managers are collectively skilled at timing the market or high-active-ownership industries. In other words, these fund managers are informed about market- or industry-wide prospects, so that they incorporate news into the prices of high-AO stocks before the rest of the market reacts to such news. Notice that the channel is a statement about the entire active mutual fund sector in aggregate. It could be that no individual fund possesses a sufficiently accurate market-timing signal to profit net of trading costs, i.e., only the combined wisdom of all, or at least many, managers suffices to effectively forecast market performance. Notice also that our explanation does not require that active managers make money trading with other investors; indeed, no such trading is required at all. Even if markets were completely segmented, so that some stocks were traded only by active mutual funds and some only by others, it could be the case that active managers have quality market signals. Those signals could be observed via their impact on the prices of the high-AO stocks: as managers trade among themselves, good news will show up in higher prices of these assets. The owners of the other stocks could then observe those price signals and push prices of low-AO stocks in the same direction, creating the lead-lag relation.

Several empirical observations support our channel. First, we can only extract predictive signals from stocks heavily owned by active mutual funds or investment advisors, not from those primarily held by other types of financial institutions such as banks, insurance companies, or pension funds. This pattern is consistent with our channel, as these alternative financial institutions are not in the business of making high-frequency bets to exploit their potential information advantages, so we would not expect high-frequency predictability of the market in the returns of the stocks they hold. Moreover, within the active mutual fund industry, we identify those funds with better past performances, higher trading volume, and more pronounced active traits as the source of the predictive signal. These observations further strengthen the support for the channel, because consistent with the channel, we indeed expect the more skilled fund managers to incorporate their private information better, and since our signal is predictive at a daily frequency, it should also be mostly generated by the funds engaging in active trading.

Having established the plausibility of our channel, we also produce several pieces of evidence that can distinguish our channel from closely related but slightly different competing hypotheses. We mainly focus on two alternative explanations of our findings: prediction by informed fund flows and prediction by temporary price pressure.

The transactions made by a fund are determined jointly by its managers and its investors, with the latter influencing the trades of securities via fund flows. So the price adjustments of the high-AO stocks may be caused by informed fund flows rather than managers' opinions about future market movements. To disentangle these two channels, we compared the signal extracted from the stocks held by active funds versus those from passive vehicles such as passive funds or ETFs. Empirically, we only observed signal predictability associated with the active funds but not with the passive institutions. Moreover, we replicated and extended Edelen and Warner (2001) to show that the daily aggregate mutual fund flow does not predict future market returns. Therefore, based on these additional findings, our empirical observations are more likely to reflect the information advantage of the fund managers rather than the fund investors.

Another competing hypothesis is that the high-frequency predictability of the market return is caused by the temporary price pressure exerted by the active mutual fund industry. Good (bad) fund performance generates inflows (outflows), which then cause more buying (selling) of the fund and temporarily push up (down) the security prices. However, this temporary price pressure channel is inconsistent with our intraday analysis: we find only the intraday component of our signal, not the overnight component, has predictive power for the market. The flow pressure to the funds would induce them to trade after the market open as well, once they realized good performances overnight, so the overnight component of the signal should have predictive power too. Furthermore, we don't observe any reversal in market price following the good or bad market return predicted by our signal, which is also inconsistent with the price pressure channel. In addition, although the temporary price pressure also generates a momentum effect for high active ownership stocks, it does not cause the lead-lag relation between the high and low active ownership stocks.

These additional tests are consistent with the hypothesis that the one-day market return predictability from high-AO stocks is driven by the collective information advantage of active mutual fund managers, rather than an effect caused by informed fund flows or temporary price pressure.

Other questions of interest concern the source and type of information the active managers are employing, and the mechanism by which that information becomes incorporated into prices. For example, what is the extent of the information about the individual stocks the managers hold and study, and to what extent is it about the market as a whole? Is the information primarily a result of managers skillfully interpreting public announcements on their own, or it it more a matter of a "grapevine" by which thoughtful opinions and analysis are spread? Should we imagine managers changing their opinion on what to buy and sell based on the new information, or merely pushing some trades forward in time while delaying others? While the results in this paper provide tantalizing clues on some of these issues, our current data and analytics are insufficient to answer them conclusively, so they must await future research.

# 5. Conclusion

This paper documents a new anomaly in the pricing of the US stock market. We show that the difference in performance between high-AO and low-AO stocks significantly predicts the next-day market return. The mispricing is modest each day but is consistently present day after day, so that a trading strategy built to exploit the anomaly has impressive performance. We verify that our finding is robust to various market proxies, including stock futures, the market ETF, and the spot market. Moreover, our new finding is entirely tradable; it is not confined to the US, but is prevalent all around the world.

Our evidence suggests that active investment managers, such as those who run mutual funds, have better-than-market information about the stocks that they focus on. They hold and trade before other investors, and consequently, the collective wisdom of all of the active managers gives rise to a signal that predicts the overall market in the next day or two.

We also ran various additional tests to support our explanation relative to several alternative hypotheses, including the informed fund flow channel or the temporary price pressure effect.

Overall, these findings have significant implications for our understanding of market efficiency and the role of professional investors in pushing prices toward fair value. With regard to stock market efficiency, we show that the market is indeed predictable one day ahead. And contrary to conventional wisdom, a simple market-timing strategy that exploits such an effect does generate highly profitable performance. Lastly, our findings also serve as a strong testament to the competence of the active mutual fund sector as a group. Even though the whole industry only generates a modest pre-fee alpha on average, our findings suggest that the active mutual funds might play a more important role in improving the information efficiency of security prices. So in sum, our findings imply that the stock market might be less efficient, and the active mutual funds more informed, than what common beliefs would suggest.

# References

Appel, Ian R, Todd A Gormley, and Donald B Keim, 2016, Passive investors, not passive owners, *Journal of Financial Economics* 121, 111–141.

Baltussen, Guido, Sjoerd van Bekkum, and Zhi Da, 2019, Indexing and stock market serial dependence around the world, *Journal of Financial Economics* 132, 26–48.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, 2012, Hedge fund stock trading in the financial crisis of 2007–2009, *The Review of Financial Studies* 25, 1–54.

Ben-David, Itzhak, Francesco Franzoni, Rabih Moussawi, and John Sedunov, 2021, The granular nature of large institutional investors, *Management Science* .

Bennett, James A, Richard W Sias, and Laura T Starks, 2003, Greener pastures and the impact of dynamic institutional preferences, *The Review of Financial Studies* 16, 1203–1238.

Bogousslavsky, Vincent, 2021, The cross-section of intraday and overnight returns, *Journal of Financial Economics* .

Bollen, Nicolas PB, and Jeffrey A Busse, 2001, On the timing ability of mutual fund managers, *The Journal of Finance* 56, 1075–1094.

Bollerslev, Tim, George Tauchen, and Hao Zhou, 2009, Expected stock returns and variance risk premia, *The Review of Financial Studies* 22, 4463–4492.

Campbell, John Y, Sanford J Grossman, and Jiang Wang, 1993, Trading volume and serial correlation in stock returns, *The Quarterly Journal of Economics* 108, 905–939.

Campbell, John Y, and Samuel B Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *The Review of Financial Studies* 21, 1509–1531.

Cao, Charles, Timothy T Simin, and Ying Wang, 2013, Do mutual fund managers time market liquidity?, *Journal of Financial Markets* 16, 279–307.

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

Chen, Yixin, Randolph B Cohen, and Zixuan Kevin Wang, 2020, Famous firms, earnings clusters, and the stock market, *Earnings Clusters, and the Stock Market (September 2, 2020)* .

Clark, Todd E, and Kenneth D West, 2007, Approximately normal tests for equal predictive accuracy in nested models, *Journal of econometrics* 138, 291–311.

Cremers, KJ Martijn, and Antti Petajisto, 2009, How active is your fund manager? a new measure that predicts performance, *The review of financial studies* 22, 3329–3365.

Cremers, Martijn, Miguel A Ferreira, Pedro Matos, and Laura Starks, 2016, Indexing and active fund management: International evidence, *Journal of Financial Economics* 120, 539–560.

Cremers, Martijn, and Ankur Pareek, 2016, Patient capital outperformance: The investment skill of high active share managers who trade infrequently, *Journal of Financial Economics* 122, 288–306.

Daniel, Kent, David Hirshleifer, and Lin Sun, 2020, Short-and long-horizon behavioral factors, *The review of financial studies* 33, 1673–1736.

Dong, Xi, Yan Li, David Rapach, and Guofu Zhou, 2021, Anomalies and the expected market return, *Baruch College Zicklin School of Business Research Paper* 02.

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

Engelberg, Joseph, R David McLean, Jeffrey Pontiff, and Matthew Ringgenberg, 2019, Are cross-sectional predictors good market-level predictors?, in *American Finance Association Annual Meeting Paper*.

Evans, Richard B, 2010, Mutual fund incubation, *The Journal of Finance* 65, 1581–1611.

Falkenstein, Eric G, 1996, Preferences for stock characteristics as revealed by mutual fund portfolio holdings, *The Journal of Finance* 51, 111–135.

Fama, Eugene F, and Kenneth R French, 2008, Dissecting anomalies, *The Journal of Finance* 63, 1653–1678.

Fama, Eugene F, and Kenneth R French, 2010, Luck versus skill in the cross-section of mutual fund returns, *The Journal of Finance* 261–300.

Fama, Eugene F, and Kenneth R French, 2015, A five-factor asset pricing model, *Journal of financial economics* 116, 1–22.

Fang, Lily H, Joel Peress, and Lu Zheng, 2014, Does media coverage of stocks affect mutual funds' trading and performance?, *The Review of Financial Studies* 27, 3441–3466.

French, Kenneth R, 2008, Presidential address: The cost of active investing, *The Journal of Finance* 63, 1537–1573.

Gao, Lei, Yufeng Han, Sophia Zhengzi Li, and Guofu Zhou, 2018, Market intraday momentum, *Journal of Financial Economics* 129, 394–414.

Greenwood, Robin, and Samuel G Hanson, 2012, Share issuance and factor timing, *The Journal of Finance* 67, 761–798.

Greenwood, Robin, and Samuel G Hanson, 2013, Issuer quality and corporate bond returns, *The Review of Financial Studies* 26, 1483–1525.

Grossman, Sanford J, and Joseph E Stiglitz, 1980, On the impossibility of informationally efficient markets, *The American economic review* 70, 393–408.

Hameed, Allaudeen, Matthijs Lof, and Matti Suominen, 2017, Slow trading and stock return predictability, *Working Paper* .

Henriksson, Roy D, and Robert C Merton, 1981, On market timing and investment performance. ii. statistical procedures for evaluating forecasting skills, *Journal of business* 513–533.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang, 2019, Which factors?, *Review of Finance* 23, 1–35.

Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating anomalies, *The Review of Financial Studies* 33, 2019–2133.

Huang, Dashan, Fuwei Jiang, Jun Tu, and Guofu Zhou, 2015, Investor sentiment aligned: A powerful predictor of stock returns, *The Review of Financial Studies* 28, 791–837.

Jiang, Fuwei, Joshua Lee, Xiumin Martin, and Guofu Zhou, 2019, Manager sentiment and stock returns, *Journal of Financial Economics* 132, 126–149.

Jiang, George J, Tong Yao, and Tong Yu, 2007, Do mutual funds time the market? evidence from portfolio holdings, *Journal of Financial Economics* 86, 724–758.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, 2014, Time-varying fund manager skill, *The Journal of Finance* 69, 1455–1484.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, 2005, On the industry concentration of actively managed equity mutual funds, *The Journal of Finance* 60, 1983–2011.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, 2008, Unobserved actions of mutual funds, *The Review of Financial Studies* 21, 2379–2416.

Kelly, Bryan, and Seth Pruitt, 2013, Market expectations in the cross-section of present values, *The Journal of Finance* 68, 1721–1756.

Koijen, Ralph SJ, Tobias J Moskowitz, Lasse Heje Pedersen, and Evert B Vrugt, 2018, Carry, *Journal of Financial Economics* 127, 197–225.

Koijen, Ralph SJ, and Motohiro Yogo, 2019, A demand system approach to asset pricing, *Journal of Political Economy* 127, 1475–1515.

Kyle, Albert S, 1985, Continuous auctions and insider trading, *Econometrica: Journal of the Econometric Society* 1315–1335.

Lakonishok, Josef, and Seymour Smidt, 1988, Are seasonal anomalies real? a ninety-year perspective, *The review of financial studies* 1, 403–425.

Lewellen, Jonathan, 2014, The cross section of expected stock returns, *Forthcoming in Critical Finance Review, Tuck School of Business Working Paper* .

Lo, Andrew W, and A Craig MacKinlay, 1990, When are contrarian profits due to stock market overreaction?, *The review of financial studies* 3, 175–205.

Lucca, David O, and Emanuel Moench, 2015, The pre-fomc announcement drift, *The Journal of finance* 70, 329–371.

Malkiel, Burton G, 1995, Returns from investing in equity mutual funds 1971 to 1991, *The Journal of finance* 50, 549–572.

Martin, Ian, 2017, What is the expected return on the market?, *The Quarterly Journal of Economics* 132, 367–433.

Massa, Massimo, Ludovic Phalippou, et al., 2004, Mutual funds and the market for liquidity, Technical report, CEPR Discussion Papers.

McLean, R David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability?, *The Journal of Finance* 71, 5–32.

Moskowitz, Tobias J, Yao Hua Ooi, and Lasse Heje Pedersen, 2012, Time series momentum, *Journal of financial economics* 104, 228–250.

Newey, Whitney K, and Kenneth D West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Pástor, L'uboš, Robert F Stambaugh, and Lucian A Taylor, 2017, Do funds make more when they trade more?, *The Journal of Finance* 72, 1483–1528.

Rapach, David E, Matthew C Ringgenberg, and Guofu Zhou, 2016, Short interest and aggregate stock returns, *Journal of Financial Economics* 121, 46–65.

Ross, Steve, 2015, The recovery theorem, *The Journal of Finance* 70, 615–648.

Savor, Pavel, and Mungo Wilson, 2013, How much do investors care about macroeconomic risk? evidence from scheduled economic announcements, *Journal of Financial and Quantitative Analysis* 48, 343–375.

Schumacher, David, 2018, Home bias abroad: Domestic industries and foreign portfolio choice, *The Review of Financial Studies* 31, 1654–1706.

Solomon, David H, Eugene Soltes, and Denis Sosyura, 2014, Winners in the spotlight: Media coverage of fund holdings as a driver of flows, *Journal of Financial Economics* 113, 53–72.

Wang, Jiang, 1993, A model of intertemporal asset prices under asymmetric information, *The Review of Economic Studies* 60, 249–282.

Welch, Ivo, and Amit Goyal, 2008, A comprehensive look at the empirical performance of equity premium prediction, *The Review of Financial Studies* 21, 1455–1508.

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

Zhu, Qifei, 2020, The missing new funds, *Management Science* 66, 1193–1204.

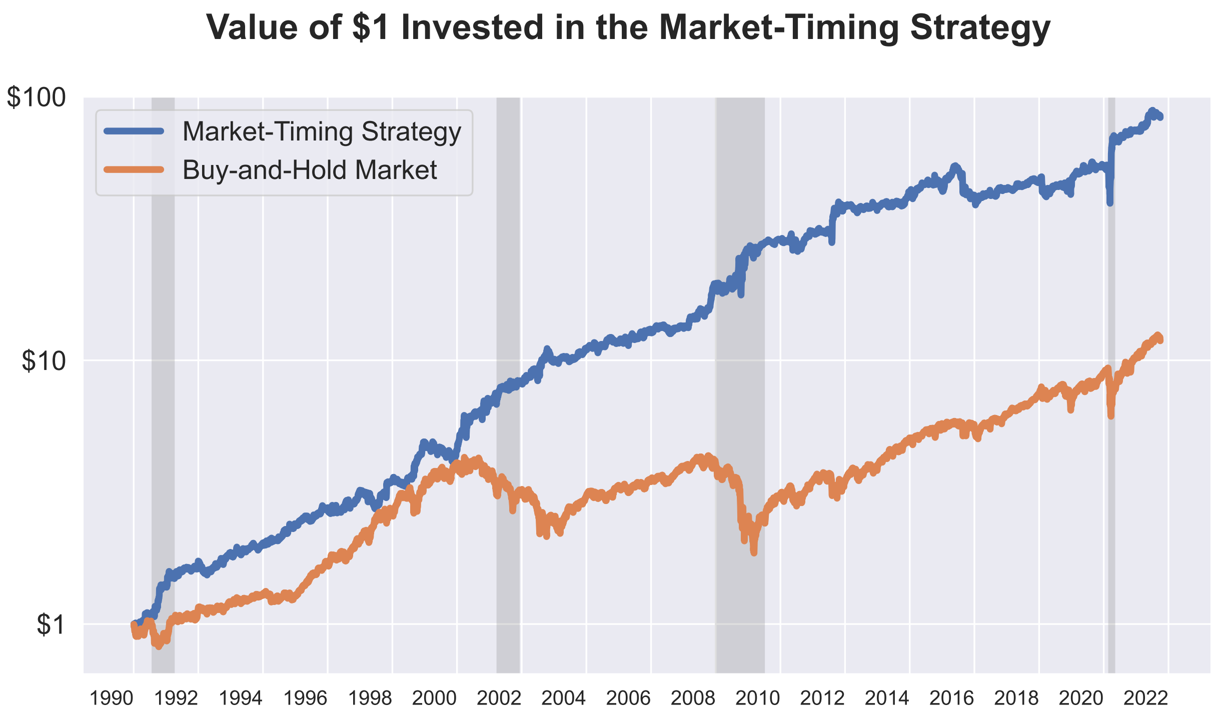


Fig. 1. **Performance of the Daily Market-Timing Strategy**

The figure plots the (log) cumulative performance of our daily market-timing strategy trading the S&P 500 futures. Following Campbell and Thompson (2008) and Gao et al. (2018), our market-timing strategy is based on the optimal portfolio for a mean-variance investor with a risk aversion coefficient of 5, using our active-ownership signal. Specifically, our strategy adjusts the weight on the S&P 500 futures using the following formula:

$$w_t = \frac{\widehat{E}_t(r^e_{m,t+1})}{5 \times \widehat{Var}_t(r^e_{m,t+1})},$$

where $\widehat{E}_t(r^e_{m,t+1})$ is the out-of-sample expected return of the S&P 500 futures estimated from the predictive regression in Table 1 using the data of July 1982 to the date of portfolio formation and $\widehat{Var}_t(r^e_{m,t+1})$ is the out-of-sample variance estimated with a rolling window of 252 trading days. The weight is bounded between -0.5 and 1.5. The sample period is from July 1982 to September 2021, and portfolio formation starts in January 1990. The blue line is the cumulative performance of the trading strategy; the orange line is the cumulative performance of the S&P 500 futures. The sample period is from July 1982 to September 2021, and portfolio formation starts in January 1990. The shaded areas denote the NBER recessions.

Table 1: **Daily S&P 500 Futures Return Predictability**

This table documents the predictability of the daily S&P 500 futures return by the lagged active-ownership signal:

$$r^e_{m,t+1} = a_0 + a_1 f(s_t) + a_2 r^e_{m,t} + \epsilon_{t+1},$$

where $r^e_{m,t+1}$ is the S&P 500 futures return on date $t+1$; $s_t$ is the active-ownership signal on date $t$, defined as the difference between the equal-weighted average returns of the high-active-ownership stocks and low-active-ownership stocks; and $f(s_t)$ is a monotonic transformation of $s_t$, including: $s_t$ itself, its sign ($Sign(s_t)$), its positive component $s^+_t$ ($\equiv \max(s_t, 0)$), and its negative component $s^-_t$ ($\equiv \min(s_t, 0)$). $\{s_t\}$ is extracted from the universe of all-but-micro-cap stocks, defined as the stocks with market capitalization above the 20th percentile of NYSE stocks. Newey and West (1987) $t$-statistics are reported in brackets. Coefficients that are significant at the 5% confidence level are in bold. $R^2$ is the in-sample R-squared from the predictive regression, and $R^2_{OOS}$ is the Campbell and Thompson (2008) out-of-sample R-squared, whose statistical significance is based on $p$-value of the Clark and West (2007) statistic, and "***" indicates significance at the 1% level. The sample period is from July 1982 to September 2021.

| | $r^e_{m,t+1}$ | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $s_t$ | **0.119** | **0.210** | | | |
| | [3.77] | [4.83] | | | |
| $Sign(s_t)$ | | | **0.064** | | |
| | | | [4.42] | | |
| $s^+_t$ | | | | **0.276** | |
| | | | | [3.94] | |
| $s^-_t$ | | | | | **0.258** |
| | | | | | [4.43] |
| $r^e_{m,t}$ | | **-0.094** | **-0.079** | **-0.084** | **-0.082** |
| | | [-4.11] | [-3.69] | [-3.75] | [-3.84] |
| N | 9896 | 9896 | 9896 | 9896 | 9896 |
| $R^2$ (%) | 0.201 | 0.957 | 0.651 | 0.816 | 0.689 |
| $R^2_{OOS}$ (%) | $0.187^{***}$ | $0.893^{***}$ | $0.641^{***}$ | $0.719^{***}$ | $0.467^{***}$ |

Table 2: **International Evidence**

This table documents the predictability of daily futures return by the lagged active-ownership signal for major equity markets around the world:

$$r_{m,t+1}^{market,e} = a_0 + a_1 s_t^{market} + a_2 r_{m,t}^{market,e} + \epsilon_{t+1},$$

where $r_{m,t+1}^{market,e}$ is the futures return of one of the ten largest equity markets: S&P 500 (United States, US), CSI 300 (China, CN), TOPIX (Japan, JP), FTSE 100 (Great Britain, GB), TSX (Canada, CA), CAC 40 (France, FR), DAX (German, DE), SMI (Switzerland, CH), NIFTY 50 (India, IN), and HSI (Hong Kong, HK). $s_t^{market}$ is the difference between the equal-weighted average returns of the high-active-ownership stocks and low-active-ownership stocks within the same market. Newey and West (1987) $t$-statistics are reported in brackets. Coefficients that are significant at the 5% confidence level are in bold. The markets are ranked in descending order based on the sample average of total market value (MV). Sample periods depend on data availability: 2000-2020 for US, JP, GB, HK, FR, DE, and IN; 2010-2020 for CN; 2009-2020 for CA; and 2003-2020 for CH.

| Equity market | $r_{m,t+1}^{US,e}$ | $r_{m,t+1}^{CN,e}$ | $r_{m,t+1}^{JP,e}$ | $r_{m,t+1}^{GB,e}$ | $r_{m,t+1}^{CA,e}$ |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $s_t^{market}$ | **0.106** | **0.508** | **0.212** | **0.186** | **0.273** |
| | [2.71] | [3.42] | [3.05] | [2.03] | [3.14] |
| $r_{m,t}^{market,e}$ | **-0.104** | 0.026 | **-0.141** | -0.026 | -0.041 |
| | [-3.38] | [0.82] | [-4.68] | [-1.37] | [-0.69] |
| N | 5153 | 2266 | 4052 | 4626 | 2625 |
| $R^2$ (%) | 0.934 | 0.971 | 1.426 | 0.069 | 0.589 |
| MV ($\$10^{12}$) | 16.31 | 5.02 | 4.11 | 2.47 | 1.95 |

| Equity market | $r_{m,t+1}^{FR,e}$ | $r_{m,t+1}^{DE,e}$ | $r_{m,t+1}^{CH,e}$ | $r_{m,t+1}^{IN,e}$ | $r_{m,t+1}^{HK,e}$ |
|---|---|---|---|---|---|
| | (6) | (7) | (8) | (9) | (10) |
| $s_t^{market}$ | **0.211** | -0.029 | **0.603** | -0.006 | 0.029 |
| | [2.83] | [-0.46] | [2.49] | [-0.12] | [0.47] |
| $r_{m,t}^{market,e}$ | **-0.063** | 0.016 | **-0.389** | -0.034 | **-0.065** |
| | [-2.53] | [0.61] | [-7.21] | [-1.35] | [-2.33] |
| N | 4378 | 4585 | 3944 | 4479 | 4196 |
| $R^2$ (%) | 0.207 | -0.032 | 1.410 | 0.069 | 0.306 |
| MV ($\$10^{12}$) | 1.83 | 1.49 | 0.96 | 0.95 | 0.85 |

Table 3: **Performance of the Daily Market-Timing Strategy**

This table evaluates the performance of our daily out-of-sample market-timing strategy for trading the S&P 500 futures. Following Campbell and Thompson (2008) and Gao et al. (2018), our market-timing strategy is based on the optimal portfolio for a mean-variance investor with a risk aversion coefficient of 5, using our active-ownership signal. Specifically, our strategy adjusts the weight on the S&P 500 futures using the following formula:

$$w_t = \frac{\widehat{E}_t(r^e_{m,t+1})}{5 \times \widehat{Var}_t(r^e_{m,t+1})},$$

where $\widehat{E}_t(r^e_{m,t+1})$ is the out-of-sample expected return of the S&P 500 futures estimated from the predictive regression in Table 1 using the data from July 1982 to the date of portfolio formation and $\widehat{Var}_t(r^e_{m,t+1})$ is the out-of-sample variance estimated with a rolling window of 252 trading days. The weight is bounded between -0.5 and 1.5. The sample period is from July 1982 to September 2021, and portfolio formation starts in January 1990. Panel A presents the key statistics (i.e., $E(r^e_t)$ and Sharpe Ratio) and the certainty equivalent return (CER) gain of the trading strategy. The CER gain is the difference between the CER for an investor who uses the predictive regression forecast of the S&P 500 futures return and the CER for an investor who uses the historical average forecast, and it can be interpreted as the management fee per annum that the investor is willing to pay so as to be indifferent between investing in the market-timing strategy with the active-ownership signal versus an alternative market-timing strategy which estimates the out-of-sample equity premium with the in-sample average. Panel B evaluates the performance of the market-timing strategy against various benchmarks, including: CAPM, Carhart 4 factors (Carhart, 1997), Fama-French 5 factors (Fama and French, 2015), Q5 (Hou et al., 2019), and Daniel-Hirshleifer-Sun behavioral factors (Daniel et al., 2020). White (1980) Heteroskedasticity-robust $t$-statistics are reported in brackets. Coefficients that are significant at the 5% confidence level are in bold. "IR" denotes the annualized information ratios relative to the benchmarks.

Panel A. Performance of the Out-of-Sample Market-Timing Strategy

| $E(r_t^e)$ (%) | Std Dev (%) | Sharpe Ratio | Skewness | Kurtosis | CER (%) |
|---|---|---|---|---|---|
| 15.26 | 16.33 | 0.95 | 0.75 | 38.45 | 5.49 |

Panel B. Regression of the Strategy's Excess Return on Alternative Factors

| | CAPM | Carhart | FF5 | Q5 | DHS |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $\alpha$ (%) | **12.21** | **11.81** | **11.92** | **12.06** | **12.42** |
| | [4.48] | [4.28] | [4.34] | [4.37] | [4.38] |
| $\beta_{mkt}$ | **0.33** | **0.34** | **0.34** | **0.35** | **0.31** |
| | [9.15] | [9.20] | [8.64] | [8.73] | [8.08] |
| $\beta_{smb}$ | | **-0.13** | **-0.11** | | |
| | | [-3.72] | [-3.31] | | |
| $\beta_{hml}$ | | 0.04 | 0.01 | | |
| | | [0.99] | [0.23] | | |
| $\beta_{umd}$ | | **0.07** | | | |
| | | [2.82] | | | |
| $\beta_{rmw}$ | | | 0.06 | | |
| | | | [1.41] | | |
| $\beta_{cma}$ | | | 0.03 | | |
| | | | [0.44] | | |
| $\beta_{r\_me}$ | | | | **-0.12** | |
| | | | | [-3.26] | |
| $\beta_{r\_ia}$ | | | | 0.02 | |
| | | | | [0.49] | |
| $\beta_{r\_roe}$ | | | | **0.14** | |
| | | | | [3.22] | |
| $\beta_{r\_eg}$ | | | | -0.07 | |
| | | | | [-1.33] | |
| $\beta_{mgmt}$ | | | | | 0.00 |
| | | | | | [0.03] |
| $\beta_{perf}$ | | | | | 0.05 |
| | | | | | [1.68] |
| IR | 0.80 | 0.78 | 0.79 | 0.79 | 0.85 |

Table 4: **Horizon of the S&P 500 Futures Return Predictability**

This table extends the main regression of Table 1 and studies the horizon of our signal's predictability of the S&P 500 futures return. Panel A shows the predictability in the following five trading days after the signal is generated. Panel B shows the predictability of the intraday and overnight components of the daily S&P 500 futures return. Following Bogousslavsky (2021), we take the price of E-mini S&P 500 futures at 9:45 am as the open price to mitigate potential microstructure issues; for the individual stock prices which we use to generate the signal, open price is defined as the average of bid and ask quotes at 9:45 am. In Panel B, "ctc," "cto," and "otc" stand for "close-to-close," "close-to-open," and "open-to-close," respectively. $\{s_t\}$ is extracted from the universe of all-but-micro-cap stocks, defined as the stocks with market capitalization above the 20th percentile of NYSE stocks. Newey and West (1987) $t$-statistics are reported in brackets. Coefficients that are significant at the 5% confidence level are in bold. The sample period in Panel A is from July 1982 to September 2021; the sample period in Panel B is from January 1998 to December 2020, during which we have the intraday data on E-mini S&P 500 futures.

Panel A. Five-Day Market Predictability

| | $r^e_{m,t+1}$ | $r^e_{m,t+2}$ | $r^e_{m,t+3}$ | $r^e_{m,t+4}$ | $r^e_{m,t+5}$ |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $s_t$ | **0.210** | 0.031 | 0.029 | -0.004 | 0.008 |
| | [4.83] | [0.63] | [0.90] | [-0.11] | [0.23] |
| $r^e_{m,t}$ | **-0.094** | **-0.065** | **-0.064** | **-0.064** | **-0.064** |
| | [-4.11] | [-3.15] | [-3.16] | [-3.17] | [-3.17] |
| N | 9896 | 9895 | 9894 | 9893 | 9892 |
| $R^2$ (%) | 0.957 | 0.409 | 0.406 | 0.394 | 0.395 |

Panel B. Intraday v.s. Overnight Market Predictability

| | $r^{ctc,e}_{m,t+1}$ | | | $r^{cto,e}_{m,t+1}$ | | | $r^{otc,e}_{m,t+1}$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| $s^{ctc}_t$ | **0.238** | | | **0.092** | | | **0.146** | | |
| | [3.75] | | | [2.19] | | | [3.26] | | |
| $s^{cto}_t$ | | 0.153 | | | 0.034 | | | -0.105 | |
| | | [1.61] | | | [0.57] | | | [-1.62] | |
| $s^{otc}_t$ | | | **0.253** | | | **0.105** | | | **0.147** |
| | | | [3.37] | | | [2.47] | | | [2.55] |
| $r^e_{m,t}$ | **-0.125** | **-0.105** | **-0.116** | **-0.045** | **-0.037** | **-0.043** | **-0.079** | **-0.063** | **-0.073** |
| | [-3.57] | [-3.10] | [-3.53] | [-2.51] | [-2.17] | [-2.61] | [-3.64] | [-2.97] | [-3.43] |
| N | 5777 | 5777 | 5777 | 5777 | 5777 | 5777 | 5777 | 5777 | 5777 |
| $R^2$ (%) | 1.647 | 1.051 | 1.508 | 0.656 | 0.358 | 0.633 | 1.067 | 0.714 | 0.945 |

Table 5: **Relation between Firm Characteristics and Active Ownership**

This table reports the time-series average of firm characteristics across stocks, sorted by active-mutual-fund ownership. The universe is limited to the stocks with market capitalization above the 20th percentile of NYSE stocks. The market capitalization, bid-ask spread, and Amihud illiquidity are calculated using CRSP stock data from July 1982 to September 2021. Analyst coverage is measured as the number of analysts releasing earnings forecasts for the firm; the data for analyst earnings forecasts are from the Institutional Brokers' Estimate System (IBES) database since January 1987. Media coverage is the number of news articles about the firm released by media outlets owned by Dow Jones & Company (e.g., *The Wall Street Journal*) in the quarter prior to portfolio formation; the data on news articles are from the RavenPack News Analytics database since January 2001.

| Active Ownership Quintile | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Active Fund Ownership (%) | 3.00 | 7.77 | 11.76 | 16.17 | 24.40 |
| Market Capitalization ($\$10^9$) | 6.23 | 9.22 | 7.37 | 4.46 | 2.96 |
| Bid-Ask Spread (%) | 1.32 | 1.00 | 0.92 | 0.88 | 0.82 |
| Amihud Illiquidity | 0.20 | 0.05 | 0.03 | 0.03 | 0.02 |
| Analyst Coverage | 8 | 11 | 12 | 11 | 11 |
| Media Coverage | 53 | 72 | 62 | 49 | 38 |

Table 6: **Daily S&P 500 Futures Return Predictability by Alternative Firm Characteristics**

This table documents the predictability of the daily S&P 500 futures return by the lagged signals extracted from returns of stocks with different firm characteristics:

$$r^e_{m,t+1} = a_0 + a_1 s_t + a_1 s^{characteristic}_t + a_2 r^e_{m,t} + \epsilon_{t+1},$$

where $r^e_{m,t+1}$ is the S&P 500 futures return on date $t+1$; $s_t$ is the difference between the equal-weighted average returns of the high-active-ownership stocks and the low-active-ownership stocks on date $t$; and $s^{characteristic}_t$ is the signal extracted from portfolios sorted by an alternative characteristic on date $t$. The signals are extracted from the universe of the all-but-micro-cap stocks, defined as the stocks with market caps above the 20th percentile of NYSE stocks. Analyst coverage is constructed using the IBES data and media coverage is constructed using the RavenPack data (see Table 5 for details). Newey and West (1987) $t$-statistics are reported in brackets. Coefficients that are significant at the 5% confidence level are in bold. The sample period for active fund ownership, market capitalization, bid-ask spread, and Amihud illiquidity is from July 1982 to September 2021; the sample period for analyst coverage starts in January 1987; the sample period for media coverage starts in January 2001.

| | $r^e_{m,t+1}$ | | | | | |
|---|---|---|---|---|---|---|
| $s^{characteristic}_t$ | (1) | (2) | (3) | (4) | (5) | (6) |
| Active Fund Ownership | **0.207** | **0.204** | **0.197** | **0.187** | **0.202** | **0.194** |
| | [4.81] | [4.55] | [4.50] | [3.91] | [3.27] | [2.86] |
| Market Capitalization | 0.012 | | | | | **-0.254** |
| | [0.21] | | | | | [-2.51] |
| Bid-Ask Spread | | -0.016 | | | | -0.007 |
| | | [-0.60] | | | | [-0.14] |
| Amihud Illiquidity | | | -0.032 | | | -0.150 |
| | | | [-0.63] | | | [-0.96] |
| Analyst Coverage | | | | 0.055 | | -0.082 |
| | | | | [0.74] | | [-1.05] |
| Media Coverage | | | | | 0.015 | 0.167 |
| | | | | | [0.20] | [1.54] |
| $r^e_{m,t}$ | **-0.095** | **-0.093** | **-0.097** | **-0.096** | **-0.114** | **-0.136** |
| | [-4.18] | [-3.96] | [-4.26] | [-3.93] | [-3.30] | [-3.25] |
| N | 9896 | 9768 | 9896 | 8757 | 5219 | 5219 |
| $R^2$ (%) | 0.951 | 0.947 | 0.977 | 1.011 | 1.329 | 1.758 |

Table 7: **Daily S&P 500 Futures Return Predictability by Alternative Financial Institutions**

This table documents the predictability of the daily S&P 500 futures return by the lagged signals extracted from returns of stocks held by alternative financial institutions:

$$r^e_{m,t+1} = a_0 + a_1 s^{institution}_t + a_3 r^e_{m,t} + \epsilon_{t+1},$$

where $r^e_{m,t+1}$ is the S&P 500 futures return on date $t+1$ and $s^{institution}_t$ is the difference between the equal-weighted average returns of the stocks with high and low ownership held by a specific type of financial institution on date $t$. The signals are extracted from the universe of all-but-micro-cap stocks, defined as the stocks with market capitalization above the 20th percentile of NYSE stocks. Mutual fund and 13F institutional holdings are from the Thomson Reuters holdings data. Hedge fund holdings are from the FactSet Global Ownership data. The institution classification follows Koijen and Yogo (2019). Newey and West (1987) $t$-statistics are reported in brackets. Coefficients that are significant at the 5% confidence level are in bold. The sample periods depend on data availability: 1988 - 2021 for passive funds and ETFs, 1999 - 2021 for hedge funds, and 1982 - 2021 for the rest of the institutions.

| | | | | | $r^e_{m,t+1}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $s^{institution}_t$ | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Active Mutual Fund | **0.216** | | | | | | | **0.241** | **0.244** |
| | [4.89] | | | | | | | [4.09] | [3.25] |
| Passive Fund and ETF | | 0.068 | | | | | | 0.106 | 0.135 |
| | | [1.90] | | | | | | [1.42] | [1.50] |
| Investment Advisor | | | **0.141** | | | | | 0.010 | -0.131 |
| | | | [2.58] | | | | | [0.16] | [-1.44] |
| Pension Fund | | | | 0.121 | | | | 0.130 | 0.170 |
| | | | | [1.84] | | | | [1.50] | [1.60] |
| Bank | | | | | 0.048 | | | -0.047 | -0.002 |
| | | | | | [0.96] | | | [-0.63] | [-0.02] |
| Insurance Company | | | | | | 0.070 | | **-0.274** | **-0.326** |
| | | | | | | [1.35] | | [-2.69] | [-2.42] |
| Hedge Fund | | | | | | | **0.154** | | **0.199** |
| | | | | | | | [2.50] | | [2.23] |
| $r^e_{m,t}$ | **-0.095** | **-0.081** | **-0.076** | **-0.068** | **-0.064** | **-0.069** | **-0.100** | **-0.110** | **-0.125** |
| | [-4.14] | [-3.58] | [-3.45] | [-3.39] | [-3.06] | [-3.47] | [-3.11] | [-4.32] | [-3.86] |
| N | 9832 | 8440 | 9832 | 9641 | 9832 | 9832 | 5411 | 8440 | 5411 |
| $R^2$ (%) | 0.983 | 0.689 | 0.607 | 0.620 | 0.447 | 0.459 | 0.994 | 1.452 | 1.816 |

Table 8: **Control Aggregate Mutual Fund Flows**

This table documents the (lack of) predictive power of the aggregate mutual fund flows for the next-day market return:

$$r^e_{m,t+1} = a_0 + a_1 s_t + a_2 flow_t + \epsilon_{t+1},$$

where $r^e_{m,t+1}$ is the S&P 500 futures return on date $t+1$; $s_t$ is the difference between the equal-weighted average returns of the high-active-ownership and the low-active-ownership stocks on date $t$; and $flow_t$ represents the aggregate daily flow to US equity mutual funds. We consider four measures of aggregate fund flows, depending on the fund classification: 1) flows to all US equity mutual funds ($flow^{all}_t$), 2) flows to all funds with the S&P500 as benchmark ($flow^{S\&P500}_t$), 3) flows to all actively-managed funds ($flow^{active}_t$), and 4) flows to all passively-managed funds ($flow^{passive}_t$). The daily fund flow is calculated as the growth rate of the aggregate daily total net asset minus the value-weighted daily return across mutual funds. Newey and West (1987) $t$-statistics are reported in brackets. Coefficients that that are significant at the 5% confidence level are in bold. The data on daily total net assets are from Morningstar Direct after August 2008, and the data of daily fund returns are from CRSP after December 1998. The sample period is from August 2008 to September 2021.

| | $r^e_{m,t+1}$ | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $s_t$ | **0.212** | **0.216** | **0.214** | **0.210** |
| | [2.11] | [2.15] | [2.12] | [2.10] |
| $flow^{all}_t$ | -0.012 | | | |
| | [-0.07] | | | |
| $flow^{S\&P500}_t$ | | -0.015 | | |
| | | [-0.79] | | |
| $flow^{active}_t$ | | | 0.093 | |
| | | | [0.75] | |
| $flow^{passive}_t$ | | | | -0.221 |
| | | | | [-1.54] |
| $r^e_{m,t}$ | **-0.125** | **-0.125** | **-0.124** | **-0.125** |
| | [-3.12] | [-3.14] | [-3.11] | [-3.13] |
| N | 3208 | 3208 | 3208 | 3208 |
| $R^2$ (%) | 1.670 | 1.695 | 1.685 | 1.811 |

Table 9: **Predictability of Aggregate Stock News Sentiment**

This table documents the predictability of our active-ownership signal for the next-day aggregate news sentiment (ANS), and the relation between the predicted ANS and the next-day market return. Panel A documents the predictability of the aggregate news sentiment by the lagged active-ownership signal:

$$ans_{t+1} = a_0 + a_1 s_t + a_2 ans_t + a_3 r^e_{m,t} + \epsilon_{t+1},$$

where $ans_{t+1}$ is defined as the market cap-weighted average of the RavenPack Composite Sentiment Score across all firms on date $t+1$; $s_t$ is the difference between the equal-weighted average returns of the high-active-ownership and low-active-ownership stocks on date $t$; and $r^e_{m,t}$ is the S&P 500 futures return on date $t$. In columns (1) and (2), ANS ($ans_{t+1}$) is measured based on all US public firms, whereas in columns (3) and (4) ANS ($ans^{S\&P500}_{t+1}$) is measured only using S&P 500 firms. The construction of ANS includes news articles released by media outlets owned by Dow Jones & Company. Panel B presents the contemporaneous regression of the S&P 500 futures return on the predicted ANS ($\widehat{ans}_{t+1}$ or $\widehat{ans}^{S\&P500}_{t+1}$), estimated from the regressions in Panel A. Newey and West (1987) $t$-statistics are reported in brackets. Coefficients that are significant at the 5% confidence level are in bold. The sample period is from January 2001 to September 2021.

Panel A. Predicting Future Aggregate News Sentiment (ANS) with Signal

| | $ans_{t+1}$ | | $ans^{S\&P500}_{t+1}$ | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $s_t$ | **5.532** | **4.698** | **5.823** | **4.907** |
| | [2.44] | [2.32] | [2.43] | [2.28] |
| $ans_t$ | | **0.446** | | |
| | | [20.56] | | |
| $ans^{S\&P500}_t$ | | | | **0.432** |
| | | | | [20.06] |
| $r^e_{m,t}$ | **2.753** | **-1.497** | **2.901** | **-1.402** |
| | [4.10] | [-2.29] | [4.01] | [-2.09] |
| N | 5155 | 5155 | 5155 | 5155 |
| $R^2$ (%) | 0.638 | 19.612 | 0.653 | 18.428 |

Panel B. Regression of Market Return on Predicted Aggregate News Sentiment (ANS)

| | $r^e_{m,t+1}$ | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $\widehat{ans}_{t+1}$ | **0.039** | **0.046** | | |
| | [3.28] | [3.20] | | |
| $ans_t$ | | **-0.021** | | |
| | | [-3.19] | | |
| $\widehat{ans}^{S\&P500}_{t+1}$ | | | **0.037** | **0.044** |
| | | | [3.28] | [3.20] |
| $ans^{S\&P500}_t$ | | | | **-0.019** |
| | | | | [-3.19] |
| $r^e_{m,t}$ | **-0.222** | -0.043 | **-0.222** | -0.051 |
| | [-3.78] | [-1.62] | [-3.78] | [-1.90] |
| N | 5155 | 5155 | 5155 | 5155 |
| $R^2$ (%) | 1.394 | 1.384 | 1.394 | 1.381 |

Table 10: **Daily S&P 500 Futures Return Predictability by Subgroups of Active Mutual Funds**

This table documents the predictability of the daily S&P 500 futures return by the lagged signals extracted from returns of stocks held by active mutual funds with different characteristics:

$$r^e_{m,t+1} = a_0 + a_1 s_t^{High\{char\}} + a_2 s_t^{Low\{char\}} + a_3 r^e_{m,t} + \epsilon_{t+1},$$

where $r^e_{m,t+1}$ is the S&P 500 futures return on date $t+1$, while $s_t^{high\{char\}}$ ($s_t^{low\{char\}}$) is the difference between the equal-weighted average date-$t$ returns of the high-ownership and low-ownership stocks held by funds in the high (low) partition according to the fund characteristic ("char"). Specifically, we consider the following fund characteristics: information ratio (IR), turnover (Turnover), return gap (RetGap), industry concentration index (ICI), and active share (AS). The information ratio of a fund is estimated from the 4-factor model Carhart (1997) in the rolling 24-month window prior to signal construction. Turnover ratio is taken directly from the CRSP Mutual Fund Database, where it is defined as the minimum of aggregated sales or purchases of securities over the past 12 months divided by the average 12-month total net assets of the fund. Return gap (Kacperczyk et al., 2008) is the difference between the actual performance of the fund and the performance of a hypothetical portfolio based on the fund's previous-quarter holdings. Industry concentration index (Kacperczyk et al., 2005) captures the extent to which the fund's portfolio is concentrated in certain industries. Active share (Cremers and Petajisto, 2009) is the sum of absolute deviations of the fund portfolio weights from the benchmark weights. The signals are extracted from the universe of all-but-micro-cap stocks, defined as the stocks with market capitalization above the 20th percentile of NYSE stocks. Newey and West (1987) $t$-statistics are reported in brackets. Coefficients that are significant at the 5% confidence level are in bold. The sample period is from July 1982 to September 2021.

| | $r^e_{m,t+1}$ | | | | |
|---|---|---|---|---|---|
| $s_t^{type}$ | (1) | (2) | (3) | (4) | (5) |
| High IR | **0.203** | | | | |
| | [4.33] | | | | |
| Low IR | 0.042 | | | | |
| | [0.94] | | | | |
| High Turnover | | **0.182** | | | |
| | | [4.95] | | | |
| Low Turnover | | 0.088 | | | |
| | | [1.84] | | | |
| High RetGap | | | **0.190** | | |
| | | | [3.64] | | |
| Low RetGap | | | 0.054 | | |
| | | | [1.14] | | |
| High ICI | | | | **0.173** | |
| | | | | [4.30] | |
| Low ICI | | | | 0.095 | |
| | | | | [1.80] | |
| High AS | | | | | **0.173** |
| | | | | | [3.32] |
| Low AS | | | | | 0.124 |
| | | | | | [1.92] |
| $r^e_{m,t}$ | **-0.093** | **-0.106** | **-0.098** | **-0.096** | **-0.097** |
| | [-4.18] | [-4.41] | [-4.38] | [-4.30] | [-4.30] |
| N | 9832 | 9578 | 9896 | 9896 | 9896 |
| $R^2$ (%) | 1.000 | 1.120 | 1.039 | 0.970 | 1.000 |

Table 11: **Daily Industry Return Predictability**

This table documents the predictability of daily industry-specific returns by the lagged active-ownership signals extracted from returns of stocks held by active mutual funds within the same industry:

$$\tilde{r}_{t+1}^{industry} = a_0 + a_1 s_t^{industry} + a_2 \tilde{r}_t^{industry} + \epsilon_t,$$

where $\tilde{r}_{t+1}^{industry} (\equiv r_{t+1}^{industry} - r_{m,t+1})$ is the value-weighted return of a specific industry in excess of the value-weighted market return on date $t+1$, while $s_t^{industry}$ is the lagged difference between the equal-weighted average returns of the high-ownership and low-ownership stocks held by active mutual funds within the industry. The signals are extracted from the universe of all-but-micro-cap stocks, defined as the stocks with market capitalization above the 20th percentile of NYSE stocks. Industry classification follows Kacperczyk et al. (2005) and includes manufacturing (manftr), business service (bussv), healthcare (hlthcr), finance (fin), wholesale (whlsl), telecom (telcm), energy (engy), non-durable (nondur), durable (dur), and utility (util). The industries are ranked in descending order based on the average active mutual fund ownership of the stocks within the industry, which is reported in the last row. Newey and West (1987) $t$-statistics are reported in brackets. Coefficients that are significant at the 5% confidence level are in bold. The sample period is from July 1982 to September 2021.

| | $\tilde{r}_{t+1}^{manftr}$ | $\tilde{r}_{t+1}^{bussv}$ | $\tilde{r}_{t+1}^{hlthcr}$ | $\tilde{r}_{t+1}^{fin}$ | $\tilde{r}_{t+1}^{whlsl}$ | $\tilde{r}_{t+1}^{telcm}$ | $\tilde{r}_{t+1}^{engy}$ | $\tilde{r}_{t+1}^{nondur}$ | $\tilde{r}_{t+1}^{dur}$ | $\tilde{r}_{t+1}^{util}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $s_t^{industry}$ | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Manufacturing | **0.032** | | | | | | | | | |
| | [2.18] | | | | | | | | | |
| Business Service | | **0.064** | | | | | | | | |
| | | [4.15] | | | | | | | | |
| Healthcare | | | **0.019** | | | | | | | |
| | | | [2.07] | | | | | | | |
| Finance | | | | **0.043** | | | | | | |
| | | | | [2.48] | | | | | | |
| Wholesale | | | | | -0.003 | | | | | |
| | | | | | [-0.20] | | | | | |
| Telecom | | | | | | -0.006 | | | | |
| | | | | | | [-0.97] | | | | |
| Energy | | | | | | | 0.018 | | | |
| | | | | | | | [0.91] | | | |
| Non-durable | | | | | | | | 0.004 | | |
| | | | | | | | | [0.42] | | |
| Durable | | | | | | | | | -0.002 | |
| | | | | | | | | | [-0.21] | |
| Utility | | | | | | | | | | -0.005 |
| | | | | | | | | | | [-0.22] |
| $\tilde{r}_t^{industry}$ | **0.121** | **0.049** | **0.073** | -0.021 | **0.043** | **0.051** | **0.047** | **0.040** | -0.008 | **0.072** |
| | [7.06] | [3.42] | [4.78] | [-0.79] | [2.69] | [3.12] | [3.11] | [2.20] | [-0.43] | [4.10] |
| N | 9896 | 9896 | 9896 | 9896 | 9896 | 9896 | 9896 | 9896 | 9896 | 9896 |
| $R^2$ (%) | 1.721 | 0.902 | 0.628 | 0.174 | 0.161 | 0.249 | 0.280 | 0.138 | -0.013 | 0.487 |
| AO (%) | 12.65 | 12.22 | 11.32 | 11.01 | 10.89 | 9.67 | 9.18 | 9.16 | 8.82 | 8.53 |

Table 12: **Cross-Predictability of Daily Stock Returns by Active Mutual Fund Ownership**

This table documents the lead-lag relation of daily returns among stocks with different levels of active mutual fund ownership. To ensure the results are tradable, we predict the open-to-close returns with the close-to-close returns from the previous trading day, skipping the overnight period. Following Bogousslavsky (2021), the open price of an individual stock is defined as the average of bid and ask quotes at 9:45am to mitigate potential microstructure issues. The universe of all-but-micro-cap stocks, defined as the stocks with market capitalization above the 20th percentile of NYSE stocks, are divided into five groups according to their previous-quarter active mutual fund ownership. Equal-weighted and value-weighted returns are computed for each group of stocks. Newey and West (1987) $t$-statistics are reported in brackets. Coefficients that that are significant at the 5% confidence level are in bold. The sample period is from October 1985 to September 2021. The intraday stock transactions data for 1985 - 1992 are from the Institute for the Study of Security Markets (ISSM) database, and the data for 1993 - 2021 are from NYSE Trade and Quote (TAQ) database.

Panel A. Equal-Weight Portfolios Formed by Active Mutual Fund Ownership

| | $r_{t+1}^{h,ew}$ | $r_{t+1}^{4,ew}$ | $r_{t+1}^{3,ew}$ | $r_{t+1}^{2,ew}$ | $r_{t+1}^{l,ew}$ |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $r_t^{h,ew}$ | **0.313** | **0.266** | **0.248** | **0.238** | **0.221** |
| | [4.69] | [4.11] | [3.84] | [3.82] | [3.67] |
| $r_t^{4,ew}$ | 0.043 | 0.071 | 0.069 | 0.052 | 0.039 |
| | [0.50] | [0.85] | [0.82] | [0.60] | [0.45] |
| $r_t^{3,ew}$ | -0.092 | -0.096 | -0.128 | -0.184 | **-0.211** |
| | [-0.93] | [-0.98] | [-1.26] | [-1.78] | [-2.01] |
| $r_t^{2,ew}$ | **-0.246** | **-0.245** | -0.200 | -0.149 | -0.118 |
| | [-2.35] | [-2.37] | [-1.93] | [-1.36] | [-1.08] |
| $r_t^{l,ew}$ | -0.069 | -0.044 | -0.030 | 0.003 | 0.039 |
| | [-1.15] | [-0.75] | [-0.51] | [0.05] | [0.71] |
| N | 9075 | 9075 | 9075 | 9075 | 9075 |
| $R^2$ (%) | 1.304 | 1.142 | 0.961 | 0.793 | 0.608 |

Panel B. Value-Weight Portfolios Formed by Active Mutual Fund Ownership

| | $r_{t+1}^{h,vw}$ | $r_{t+1}^{4,vw}$ | $r_{t+1}^{3,vw}$ | $r_{t+1}^{2,vw}$ | $r_{t+1}^{l,vw}$ |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $r_t^{h,vw}$ | **0.142** | **0.123** | **0.090** | **0.087** | **0.105** |
| | [3.41] | [3.19] | [2.39] | [2.54] | [3.23] |
| $r_t^{4,vw}$ | -0.068 | -0.079 | -0.044 | -0.060 | -0.092 |
| | [-1.00] | [-1.20] | [-0.80] | [-1.19] | [-1.84] |
| $r_t^{3,vw}$ | 0.017 | 0.021 | 0.002 | -0.016 | -0.017 |
| | [0.31] | [0.41] | [0.04] | [-0.37] | [-0.42] |
| $r_t^{2,vw}$ | -0.068 | -0.044 | -0.039 | -0.061 | -0.084 |
| | [-1.34] | [-0.86] | [-0.88] | [-1.43] | [-1.94] |
| $r_t^{l,vw}$ | **-0.075** | **-0.077** | **-0.076** | -0.014 | 0.040 |
| | [-2.01] | [-2.10] | [-2.26] | [-0.49] | [0.98] |
| N | 9075 | 9075 | 9075 | 9075 | 9075 |
| $R^2$ (%) | 0.518 | 0.544 | 0.658 | 0.696 | 0.759 |

# Appendix A. Proof of Propositions

**Proposition 1.** *The aforementioned economy features a linear equilibrium, where the price of a high-AO stock is*

$$P_{i,t}^{Hi} = \frac{1}{R-1}\left(\mu - A^{Hi}\right) + \frac{\phi}{R-\phi}\left(d_{a,t} + d_{i,t}^{Hi}\right) + B^{Hi}\left(\epsilon_{a,t+1} + \epsilon_{i,t+1}^{Hi}\right) + C^{Hi} u_{i,t}^{Hi}$$

*and the price of a low-AO stock is*

$$\begin{aligned} P_{i,t}^{Lo} = &\frac{1}{R-1}\left(\mu - A^{Lo}\right) + \frac{\phi}{R-\phi}\left(d_{a,t} + d_{i,t}^{Lo}\right) + B^{Lo}\left(\frac{\sum_i d_{i,t}^{Lo}}{N} - d_{i,t}^{Lo}\right) \\ &+ C^{Lo}\left(B^{Hi}\left(\frac{\sum_i \epsilon_{i,t+1}^{Hi}}{N} + \epsilon_{a,t+1}\right) + C^{Hi}\frac{\sum_i u_{i,t}^{Hi}}{N}\right) + D^{Lo} u_{i,t}^{Lo}; \end{aligned}$$

*for some constants:* $A^{Hi}, B^{Hi}, C^{Hi}, A^{Lo}, B^{Lo}, C^{Lo}$, *and* $D^{Lo}$.

*Proof.* Conjecture that the value function of an informed asset manager takes the form:

$$J^i\left(W_t\right) = -\exp\left(-a^I W_t - b^I\right);$$

and the value function of the uninformed investor is:

$$J^U\left(W_t\right) = -\exp\left(-a^U W_t - b^U\right)$$

for some constants: $a^I, b^I, a^U$ and $b^U$.

**Verification of the prices:**

Under equilibrium, the informed investor's demand for risky assets is

$$x^i\left(P_{i,t}^{Hi}\right) = \frac{\mathbb{E}\left(P_{i,t+1}^{Hi} + D_{i,t+1}^{Hi} | s_{i,t}\right) - RP_{i,t}^{Hi}}{a^I Var\left(P_{i,t+1}^{Hi} + D_{i,t+1}^{Hi} | s_{i,t}\right)}.$$

The next-period payoff of the high-AO stock $i$ is

$$
\begin{aligned}
& P_{i,t+1}^{Hi} + D_{i,t+1}^{Hi} \\
=& \left( \frac{1}{R-1} \left( \mu - A^{Hi} \right) + \frac{\phi}{R-\phi} \left( d_{a,t+1} + d_{i,t+1}^{Hi} \right) + B^{Hi} \left( \epsilon_{a,t+2} + \epsilon_{i,t+2}^{Hi} \right) + C^{Hi} u_{i,t+1}^{Hi} \right) + \left( \mu + d_{a,t+1} + d_{i,t+1}^{Hi} \right) \\
=& \frac{1}{R-1} \left( R\mu - A^{Hi} \right) + \frac{R}{R-\phi} \left( d_{a,t+1} + d_{i,t+1}^{Hi} \right) + B^{Hi} \left( \epsilon_{a,t+2} + \epsilon_{i,t+2}^{Hi} \right) + C^{Hi} u_{i,t+1}^{Hi} \\
=& \frac{1}{R-1} \left( R\mu - A^{Hi} \right) + \frac{R\phi}{R-\phi} \left( d_{a,t} + d_{i,t}^{Hi} \right) + \frac{R}{R-\phi} \left( \epsilon_{a,t+1} + \epsilon_{i,t+1}^{Hi} \right) + B^{Hi} \left( \epsilon_{a,t+2} + \epsilon_{i,t+2}^{Hi} \right) + C^{Hi} u_{i,t+1}^{Hi}.
\end{aligned}
$$

Therefore,

$$
\mathbb{E} \left( P_{i,t+1}^{Hi} + D_{i,t+1}^{Hi} | s_{i,t} \right) = \frac{1}{R-1} \left( R\mu - A^{Hi} \right) + \frac{R\phi}{R-\phi} \left( d_{a,t} + d_{i,t}^{Hi} \right) + \frac{R}{R-\phi} \left( \epsilon_{a,t+1} + \epsilon_{i,t+1}^{Hi} \right),
$$

$$
Var \left( P_{i,t+1}^{Hi} + D_{i,t+1}^{Hi} | s_{i,t} \right) = \left( B^{Hi} \right)^2 \left( \sigma_a^2 + \sigma_i^2 \right) + \left( C^{Hi} \right)^2 \left( \sigma_u^2 + \sigma_\eta^2 \right),
$$

and

$$
x^i \left( P_{i,t}^{Hi} \right) = \frac{\frac{1}{R-1} \left( R\mu - A^{Hi} \right) + \frac{R\phi}{R-\phi} \left( d_{a,t} + d_{i,t}^{Hi} \right) + \frac{R}{R-\phi} \left( \epsilon_{a,t+1} + \epsilon_{i,t+1}^{Hi} \right) - RP_{i,t}^{Hi}}{a^I \left( \left( B^{Hi} \right)^2 \left( \sigma_a^2 + \sigma_i^2 \right) + \left( C^{Hi} \right)^2 \left( \sigma_u^2 + \sigma_\eta^2 \right) \right)}.
$$

By market clearing,

$$
\begin{aligned}
& x^i \left( P_{i,t}^{Hi} \right) + u_{i,t} = 1 \\
\Rightarrow P_{i,t}^{Hi} =& \left( \frac{\mu}{R-1} - \frac{A^{Hi}}{R(R-1)} - \frac{a^I \left( B^{Hi} \right)^2}{R} \left( \sigma_a^2 + \sigma_i^2 \right) - \frac{a^I \left( C^{Hi} \right)^2}{R} \left( \sigma_u^2 + \sigma_\eta^2 \right) \right) \\
& + \frac{\phi}{R-\phi} \left( d_{a,t} + d_{i,t}^{Hi} \right) + \frac{1}{R-\phi} \left( \epsilon_{a,t+1} + \epsilon_{i,t+1} \right) \\
& + \frac{a^I}{R} \left( \left( B^{Hi} \right)^2 \left( \sigma_a^2 + \sigma_i^2 \right) + \left( C^{Hi} \right)^2 \left( \sigma_u^2 + \sigma_\eta^2 \right) \right) u_{i,t}^{Hi},
\end{aligned}
$$

where the parameters solve

$$
\begin{aligned}
A^{Hi} &= a^I \left( \left( B^{Hi} \right)^2 \left( \sigma_a^2 + \sigma_i^2 \right) + \left( C^{Hi} \right)^2 \left( \sigma_u^2 + \sigma_\eta^2 \right) \right) \\
B^{Hi} &= \frac{1}{R-\phi} \\
C^{Hi} &= \frac{a^I}{R} \left( \left( B^{Hi} \right)^2 \left( \sigma_a^2 + \sigma_i^2 \right) + \left( C^{Hi} \right)^2 \left( \sigma_u^2 + \sigma_\eta^2 \right) \right)
\end{aligned}
$$

Under equilibrium, the uninformed investor takes the average price of the high-AO stocks as a signal of the aggregate shock, i.e.,

$$s_{u,t} \equiv B^{Hi}\left(\frac{\sum_i \epsilon^{Hi}_{i,t+1}}{N} + \epsilon_{a,t+1}\right) + C^{Hi}\frac{\sum_i u^{Hi}_{i,t}}{N}.$$

The risky demand of the uninformed investor is

$$x^U\left(\overrightarrow{P}^{Lo}_t\right) = \frac{1}{a^U}\left[\Sigma\left(\overrightarrow{P}^{Lo}_{t+1} + \overrightarrow{D}^{Lo}_{t+1}|s_{u,t}\right)\right]^{-1}\left[\mathbb{E}\left(\overrightarrow{P}^{Lo}_{t+1} + \overrightarrow{D}^{Lo}_{t+1}|s_{u,t}\right) - R\overrightarrow{P}^{Lo}_t\right],$$

where $\overrightarrow{P}^{Lo}_t$ $(\overrightarrow{D}^{Lo}_t)$ is a $N \times 1$ vector of the prices (dividends) of the low-AO stocks; $\Sigma(\cdot)$ denotes the covariance matrix.

$$\begin{aligned}
\overrightarrow{P}^{Lo}_{t+1} + \overrightarrow{D}^{Lo}_{t+1} = {} & \frac{1}{R-1}\left(\mu - A^{Lo}\right)\overrightarrow{e} + \frac{\phi}{R-\phi}\left(d_{a,t+1}\overrightarrow{e} + \overrightarrow{d}^{Lo}_{t+1}\right) + B^{Lo}\left(\frac{\sum_i d^{Lo}_{i,t+1}}{N}\overrightarrow{e} - \overrightarrow{d}^{Lo}_{t+1}\right) \\
& + C^{Lo}\left(B^{Hi}\left(\frac{\sum_i \epsilon^{Hi}_{i,t+1}}{N} + \epsilon_{a,t+1}\right) + C^{Hi}\frac{\sum_i u^{Hi}_{i,t}}{N}\right)\overrightarrow{e} + D^{Lo}\overrightarrow{u}^{Lo}_{t+1} \\
& + \mu\overrightarrow{e} + d_{a,t+1}\overrightarrow{e} + \overrightarrow{d}^{Lo}_{t+1} \\
= {} & \frac{1}{R-1}\left(R\mu - A^{Lo}\right)\overrightarrow{e} \\
& + \phi\frac{R}{R-\phi}\left(d_{a,t}\overrightarrow{e} + \overrightarrow{d}^{Lo}_t\right) + \frac{R}{R-\phi}\left(\epsilon_{a,t+1}\overrightarrow{e} + \overrightarrow{\epsilon}^{Lo}_{t+1}\right) \\
& + \phi B^{Lo}\left(\frac{\sum_i d^{Lo}_{i,t}}{N}\overrightarrow{e} - \overrightarrow{d}^{Lo}_t\right) + B^{Lo}\left(\frac{\sum_i \epsilon^{Lo}_{i,t+1}}{N}\overrightarrow{e} - \overrightarrow{\epsilon}^{Lo}_{t+1}\right) \\
& + C^{Lo}\left(B^{Hi}\left(\frac{\sum_i \epsilon^{Hi}_{i,t+2}}{N} + \epsilon_{a,t+2}\right) + C^{Hi}\frac{\sum_i u^{Hi}_{i,t+1}}{N}\right)\overrightarrow{e} + D^{Lo}\overrightarrow{u}^{Lo}_{t+1}
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}\left(\vec{P}^{Lo}_{t+1} + \vec{D}^{Lo}_{t+1} | s_{u,t}\right) &= \frac{1}{R-1}\left(R\mu - A^{Lo}\right)\vec{e} \\
&+ \phi\frac{R}{R-\phi}\left(d_{a,t}\vec{e} + \vec{d}^{Lo}_{t}\right) + \phi B^{Lo}\left(\frac{\sum_i d^{Lo}_{i,t}}{N}\vec{e} - \vec{d}^{Lo}_{t}\right) \\
&+ \frac{R}{R-\phi}\frac{B^{Hi}\sigma_a^2}{\left(\left(B^{Hi}\right)^2\left(\frac{\sigma_i^2}{N} + \sigma_a^2\right) + \left(D^{Hi}\right)^2\left(\frac{\sigma_\eta^2}{N} + \sigma_u^2\right)\right)} s_{u,t}\cdot\vec{e} \\
&\equiv \frac{1}{R-1}\left(R\mu - A^{Lo}\right)\vec{e} \\
&+ \phi\frac{R}{R-\phi}\left(d_{a,t}\vec{e} + \vec{d}^{Lo}_{t}\right) + \phi B^{Lo}\left(\frac{\sum_i d^{Lo}_{i,t}}{N}\vec{e} - \vec{d}^{Lo}_{t}\right) \\
&+ \frac{R}{R-\phi}E^{Lo}s_{u,t}\cdot\vec{e}
\end{aligned}
$$

$$
\Sigma\left(\vec{P}^{Lo}_{t+1} + \vec{D}^{Lo}_{t+1} | s_{u,t}\right) = \begin{bmatrix} \Omega & \omega & \cdots & \omega \\ \omega & \Omega & & \\ \vdots & & \ddots & \\ \omega & \cdots & \omega & \Omega \end{bmatrix}
$$

where

$$
E^{Lo} = \frac{B^{Hi}\sigma_a^2}{\left(\left(B^{Hi}\right)^2\left(\frac{\sigma_i^2}{N} + \sigma_a^2\right) + \left(D^{Hi}\right)^2\left(\frac{\sigma_\eta^2}{N} + \sigma_u^2\right)\right)}
$$

$$
\begin{aligned}
\Omega = &\left[\left(\frac{R}{R-\phi} - \left(1 - \frac{1}{N}\right)B^{Lo}\right)^2 + (N-1)\left(\frac{B^{Lo}}{N}\right)^2\right]\sigma_i^2 + \left(\frac{R}{R-\phi}\right)^2\sigma^2_{\epsilon_a|s_u} \\
&+ \left(C^{Lo}\right)^2\left(\left(B^{Hi}\right)^2\left(\frac{\sigma_i^2}{N} + \sigma_a^2\right) + \left(C^{Hi}\right)^2\frac{\sigma_u^2}{N}\right) + \left(D^{Lo}\right)^2\left(\sigma_u^2 + \sigma_\eta^2\right)
\end{aligned}
$$

$$
\omega = \left(\frac{R}{R-\phi}\right)^2\sigma^2_{\epsilon_a|s_u} + \left(B^{Lo}\right)^2\frac{\sigma_i^2}{N} + \left(C^{Lo}\right)^2\left(\left(B^{Hi}\right)^2\left(\frac{\sigma_i^2}{N} + \sigma_a^2\right) + \left(C^{Hi}\right)^2\frac{\sigma_u^2}{N}\right) + \left(D^{Lo}\right)^2
$$

$$
\sigma^2_{\epsilon_a|s_u} = \sigma_a^2 - \left(E^{Lo}\right)^2\left(\left(B^{Hi}\right)^2\left(\frac{\sigma_i^2}{N} + \sigma_a^2\right) + \left(C^{Hi}\right)^2\left(\frac{\sigma_\eta^2}{N} + \sigma_u^2\right)\right)
$$

Since $\Sigma\left(\overrightarrow{P}^{Lo}_{t+1}+\overrightarrow{D}^{Lo}_{t+1}|s_{u,t}\right)$ is symmetric,

$$\left[\Sigma\left(\overrightarrow{P}^{Lo}_{t+1}+\overrightarrow{D}^{Lo}_{t+1}|s_{u,t}\right)\right]^{-1}=\begin{bmatrix}\Gamma & \gamma & \cdots & \gamma\\ \gamma & \Gamma & & \\ \vdots & & \ddots & \\ \gamma & \cdots & \gamma & \Gamma\end{bmatrix}$$

where

$$\begin{cases}\Gamma\Omega+(N-1)\,\gamma\omega=1\\ \gamma\Omega+\Gamma\omega+(N-2)\,\gamma\omega=0\end{cases}$$

Solve for the prices of the low-AO stocks:

$$\begin{aligned}
&x_i^U\left(\overrightarrow{P}^{Lo}_{t}\right)+u_{i,t}^{Lo}=1\\
\Rightarrow&\,(\Gamma-\gamma)\left(\frac{R\mu-A^{Lo}}{R-1}+\frac{R\phi}{R-\phi}\left(d_{a,t}+d_{i,t}^{Lo}\right)+\phi B^{Lo}\left(\frac{\sum_i d_{i,t}^{Lo}}{N}-d_{i,t}^{Lo}\right)+\frac{R}{R-\phi}E^{Lo}s_{u,t}\right)\\
&+N\gamma\left(\frac{R\mu-A^{Lo}}{R-1}+\frac{R\phi}{R-\phi}\left(d_{a,t}+d_{i,t}^{Lo}+\frac{\sum_i d_{i,t}^{Lo}}{N}-d_{i,t}^{Lo}\right)+\frac{R}{R-\phi}E^{Lo}s_{u,t}\right)\\
&-(\Gamma+(N-1)\,\gamma)\,RP_{i,t}^{Lo}+a^U u_{i,t}^{Lo}=a^U\\
\Rightarrow&P_{i,t}^{Lo}=\frac{R\mu-A^{Lo}}{R\,(R-1)}-\frac{a^U}{R\,(\Gamma+(N-1)\,\gamma)}+\frac{\phi}{R-\phi}\left(d_{a,t}+d_{i,t}^{Lo}\right)\\
&+\frac{(\Gamma-\gamma)\,\phi B^{Lo}+N\gamma\frac{R\phi}{R-\phi}}{R\,(\Gamma+(N-1)\,\gamma)}\left(\frac{\sum_i d_{i,t}^{Lo}}{N}-d_{i,t}^{Lo}\right)+\frac{1}{R-\phi}E^{Lo}s_{u,t}+\frac{a^U}{R\,(\Gamma+(N-1)\,\gamma)}u_{i,t}^{Lo}
\end{aligned}$$

where the parameters solve

$$\begin{cases}\frac{1}{R-1}\left(\mu-A^{Lo}\right)=\frac{R\mu-A^{Lo}}{R(R-1)}-\frac{a^U}{R(\Gamma+(N-1)\gamma)}\\ B^{Lo}=\frac{(\Gamma-\gamma)\phi B^{Lo}+N\gamma\frac{R\phi}{R-\phi}}{R(\Gamma+(N-1)\gamma)}\\ C^{Lo}=\frac{1}{R-\phi}E^{Lo}\\ D^{Lo}=\frac{a^U}{R(\Gamma+(N-1)\gamma)}\end{cases}$$

**Verification of the value function:**

$$
\begin{aligned}
J(W_t) &= \max_{C_t, X_t} U(C_t) + \beta E_t \left[ J(W_{t+1}) | \mathcal{F}_t \right] \\
&= \max_{C_t, X_t} U(C_t) + \beta E_t \left[ -\exp(-aW_{t+1} - b) | \mathcal{F}_t \right] \\
&= \max_{C_t} U(C_t) + \beta E_t \left[ -\exp\left( -a \left( \overrightarrow{X}_t' \cdot \left( \overrightarrow{P}_{t+1} + \overrightarrow{D}_{t+1} \right) + \left( (W_t - C_t) - \overrightarrow{X}_t' \cdot \overrightarrow{P}_t \right) R \right) - b \right) | \mathcal{F}_t \right] \\
&= \max_{C_t} - \exp(-\alpha C_t) - \beta \exp\left( -a \left( (W_t - C_t) R - c \right) - b \right) \\
&= -\exp\left( -\alpha C_t^* \right) - \beta \exp\left( -a \left( \left( W_t - C_t^* \right) R - c \right) - b \right) \\
&= -\exp\left( -\alpha \left( p + qW_t \right) \right) - \beta \exp\left( -a \left( \left( W_t - p - qW_t \right) R - c \right) - b \right) \\
&= -\exp(-\alpha p) \exp(-\alpha q W_t) - \beta \exp\left( a (pR + c) - b \right) \exp\left( -a (1 - q) W_t R \right) \\
&= -\left( \exp(-\alpha p) + \beta \exp\left( a (pR + c) - b \right) \right) \exp(-\alpha q W_t) \\
&= -\exp(-aW_t - b)
\end{aligned}
$$

where

$$
\overrightarrow{X}_t = \frac{1}{a} \left[ \Sigma \left( \overrightarrow{P}_{t+1} + \overrightarrow{D}_{t+1} | \mathcal{F}_t \right) \right]^{-1} \left[ \mathbb{E} \left( \overrightarrow{P}_{t+1} + \overrightarrow{D}_{t+1} | \mathcal{F}_t \right) - R \overrightarrow{P}_t \right]
$$

$$
c = \overrightarrow{X}_t' \mathbb{E} \left( \overrightarrow{P}_{t+1} + \overrightarrow{D}_{t+1} | \mathcal{F}_t \right) - \frac{a}{2} \overrightarrow{X}_t' \Sigma \left( \overrightarrow{P}_{t+1} + \overrightarrow{D}_{t+1} | \mathcal{F}_t \right) \overrightarrow{X}_t + \overrightarrow{X}_t' \cdot \overrightarrow{P}_t R
$$

$$
\begin{aligned}
C_t^* &= \frac{-\log \frac{a\beta R}{\alpha} + aRW_t - ac + b}{\alpha + aR} \\
&= \frac{-\log \frac{a\beta R}{\alpha} - ac + b}{\alpha + aR} + \frac{aR}{\alpha + aR} W_t \\
&\equiv p + qW_t
\end{aligned}
$$

and a, b satisfy:

$$
\begin{cases}
a = \alpha q \\
\exp(-b) = \left( \exp(-\alpha p) + \beta \exp\left( a (pR + c) - b \right) \right).
\end{cases}
.
$$

$\square$

**Proposition 2.** *The average price of the high-AO stocks gives more information about the next-period aggregate shock $\epsilon_{a,t+1}$ than the average price of the low-AO stocks, i.e.*

$$Var\left(\epsilon_{a,t+1}|P_t^{Hi}\right) < Var\left(\epsilon_{a,t+1}|P_t^{Lo}\right),$$

*where $P_t^{Hi} \equiv \frac{1}{N}\sum_i P_{i,t}^{Hi}$ and $P_t^{Lo} \equiv \frac{1}{N}\sum_i P_{i,t}^{Lo}$ are the average prices of the high-AO and low-AO stocks.*

*Proof.* The average price of the high-AO stock is

$$P_t^{Hi} = \frac{1}{R-1}\left(\mu - A^{Hi}\right) + \frac{\phi}{R-\phi}\left(d_{a,t} + \frac{\sum_i d_{i,t}^{Hi}}{N}\right) + B^{Hi}\left(\epsilon_{a,t+1} + \frac{\sum_i \epsilon_{i,t+1}^{Hi}}{N}\right) + C^{Hi}\frac{\sum_i u_{i,t}^{Hi}}{N};$$

and the average price of the low-AO stock is

$$P_t^{Lo} = \frac{1}{R-1}\left(\mu - A^{Lo}\right) + \frac{\phi}{R-\phi}\left(d_{a,t} + \frac{\sum_i d_{i,t}^{Lo}}{N}\right) + C^{Lo}\left(B^{Hi}\left(\frac{\sum_i \epsilon_{i,t+1}^{Hi}}{N} + \epsilon_{a,t+1}\right) + C^{Hi}\frac{\sum_i u_{i,t}^{Hi}}{N}\right) + D^{Lo}\frac{\sum_i u_{i,t}^{Lo}}{N}.$$

The rest is immediate from the fact that $D^{Lo} \neq 0$ and $Var\left(\frac{\sum_i u_{i,t}^{Lo}}{N}\right) > 0$. $\square$

**Proposition 3.** *Define signal $s_t$ as the return difference between the high-AO and low-AO stocks, i.e.*

$$s_t \equiv R_t^{Hi} - R_t^{Lo},$$

*where $R_t^{Hi} \equiv P_t^{Hi} + D_t^{Hi} - P_{t-1}^{Hi}$ and $R_t^{Lo} \equiv P_t^{Lo} + D_t^{Lo} - P_{t-1}^{Lo}$ are the average (dollar) returns of the high-AO and low-AO stocks.*

*Under the parameter specification where $Var\left(R_t^{Hi}\right) = Var\left(R_t^{Lo}\right)$ and $N$ is large,*

$$s_t \approx \left(1 - C^{Lo}\right) B^{Hi} \left(\epsilon_{a,t+1} - \epsilon_{a,t}\right) + \left(1 - C^{Lo}\right) C^{Hi} \left(u_t^{Hi} - u_{t-1}^{Hi}\right) + D^{Lo} \left(u_t^{Lo} - u_{t-1}^{Lo}\right)$$

*for some $C^{Lo} \in (0,1)$.*[24]

*Therefore, $s_t$ is predictive of the next-period aggregate shock $\epsilon_{a,t+1}$, and thus the next-period market return.*

*Proof.* The average (dollar) return of the high-AO stocks is

$$\begin{aligned} R_t^{Hi} &\equiv P_t^{Hi} + D_t^{Hi} - P_{t-1}^{Hi} \\ &= \mu + \frac{\phi}{R-\phi}\left(\left((R-1)\, d_{a,t-1} + \frac{R}{\phi}\epsilon_{a,t}\right) + \left((R-1)\frac{\sum_i d_{i,t-1}^{Hi}}{N} + \frac{R}{\phi}\frac{\sum_i \epsilon_{i,t}^{Hi}}{N}\right)\right) \\ &+ B^{Hi}\left(\left(\epsilon_{a,t+1} + \frac{\sum_i \epsilon_{i,t+1}^{Hi}}{N}\right) - \left(\epsilon_{a,t} + \frac{\sum_i \epsilon_{i,t}^{Hi}}{N}\right)\right) + C^{Hi}\left(\frac{\sum_i u_{i,t}^{Hi}}{N} - \frac{\sum_i u_{i,t-1}^{Hi}}{N}\right); \end{aligned}$$

the average (dollar) return of the low-AO stocks is

$$\begin{aligned} R_t^{Lo} &\equiv P_t^{Lo} + D_t^{Lo} - P_{t-1}^{Lo} \\ &= \mu + \frac{\phi}{R-\phi}\left(\left((R-1)\, d_{a,t-1} + \frac{R}{\phi}\epsilon_{a,t}\right) + \left((R-1)\frac{\sum_i d_{i,t-1}^{Lo}}{N} + \frac{R}{\phi}\frac{\sum_i \epsilon_{i,t}^{Lo}}{N}\right)\right) \\ &+ C^{Lo}B^{Hi}\left(\left(\epsilon_{a,t+1} + \frac{\sum_i \epsilon_{i,t+1}^{Hi}}{N}\right) - \left(\epsilon_{a,t} + \frac{\sum_i \epsilon_{i,t}^{Hi}}{N}\right)\right) + C^{Lo}C^{Hi}\left(\frac{\sum_i u_{i,t}^{Hi}}{N} - \frac{\sum_i u_{i,t-1}^{Hi}}{N}\right) \\ &+ D^{Lo}\left(\frac{\sum_i u_{i,t}^{Lo}}{N} - \frac{\sum_i u_{i,t-1}^{Lo}}{N}\right). \end{aligned}$$

[24]Empirically, the daily volatility of the high-AO portfolio is close to the low-AO portfolio (1.31% VS 1.15%).

The signal $s_t$ is

$$
\begin{aligned}
s_t &\equiv R_t^{Hi} - R_t^{Lo} \\
&= \frac{\phi}{R-\phi}\left((R-1)\,\frac{\sum_i d_{i,t-1}^{Hi} - \sum_i d_{i,t-1}^{Lo}}{N} + \frac{R}{\phi}\frac{\sum_i \epsilon_{i,t}^{Hi} - \sum_i \epsilon_{i,t}^{Lo}}{N}\right) \\
&+ \left(1-C^{Lo}\right) B^{Hi}\left(\left(\epsilon_{a,t+1} + \frac{\sum_i \epsilon_{i,t+1}^{Hi}}{N}\right) - \left(\epsilon_{a,t} + \frac{\sum_i \epsilon_{i,t}^{Hi}}{N}\right)\right) \\
&+ \left(1-C^{Lo}\right) C^{Hi}\left(\frac{\sum_i u_{i,t}^{Hi}}{N} - \frac{\sum_i u_{i,t-1}^{Hi}}{N}\right) \\
&+ D^{Lo}\left(\frac{\sum_i u_{i,t}^{Lo}}{N} - \frac{\sum_i u_{i,t-1}^{Lo}}{N}\right) \\
&\approx \left(1-C^{Lo}\right) B^{Hi}\left(\epsilon_{a,t+1} - \epsilon_{a,t}\right) \\
&+ \left(1-C^{Lo}\right) C^{Hi}\left(u_t^{Hi} - u_{t-1}^{Hi}\right) + D^{Lo}\left(u_t^{Lo} - u_{t-1}^{Lo}\right).
\end{aligned}
$$

If $Var\left(R_t^{Hi}\right) = Var\left(R_t^{Lo}\right)$, then $C^{Lo} < 1$. The rest is immediate. □

# Appendix B. Robustness Checks

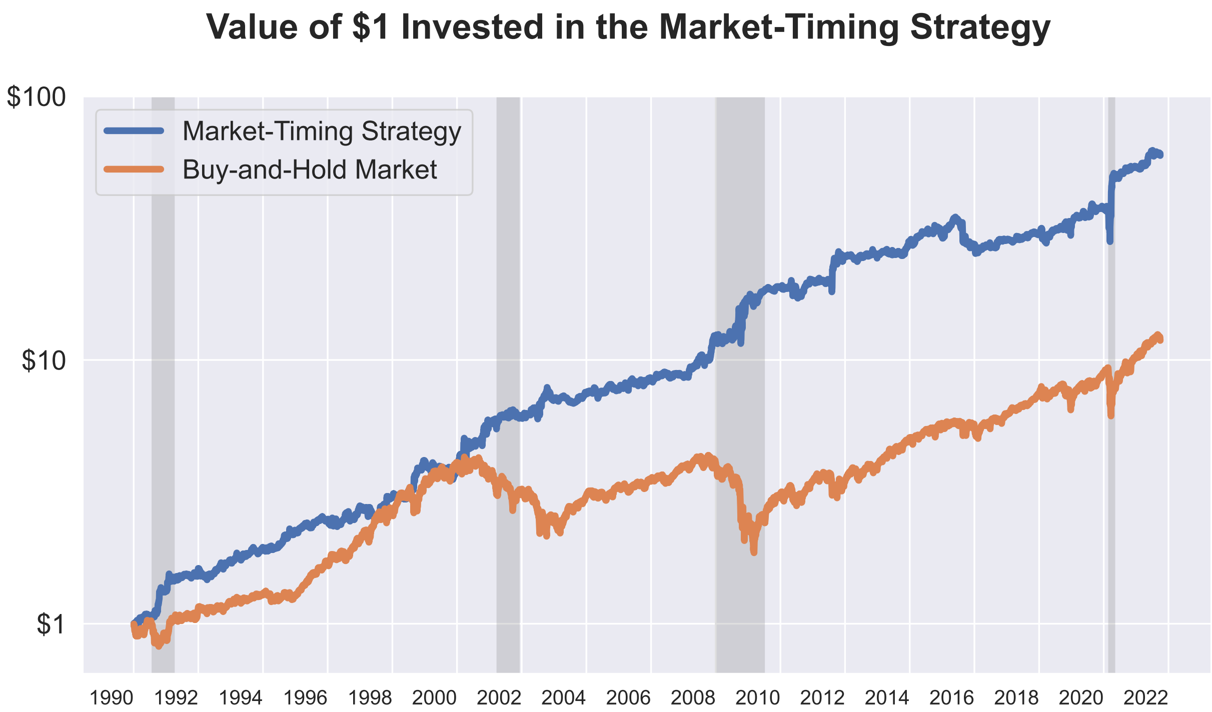


Fig. B1. **Performance of a Daily Market-Timing Strategy: Signals Extracted From the 2-Month Lagged Active Ownership**

The figure plots the (log) cumulative performance of our daily market-timing strategy for trading the S&P 500 futures. Following Campbell and Thompson (2008) and Gao et al. (2018), our market-timing strategy is based on the optimal portfolio for a mean-variance investor with a risk aversion coefficient of 5, using our active-ownership signal. Specifically, our strategy adjusts the weight on the S&P 500 futures using the following formula:

$$w_t = \frac{\widehat{E}_t(r^e_{m,t+1})}{5 \times \widehat{Var}_t(r^e_{m,t+1})},$$

where $\widehat{E}_t(r^e_{m,t+1})$ is the out-of-sample expected return of the S&P 500 futures estimated from the predictive regression in Table 1 using the data of July 1982 to the date of portfolio formation and $\widehat{Var}_t(r^e_{m,t+1})$ is the out-of-sample variance estimated with a rolling window of 252 trading days. The weight is bounded between -0.5 and 1.5. The sample period is from July 1982 to September 2021, and portfolio formation starts in January 1990. The blue line is the cumulative performance of the trading strategy; the orange line is the cumulative performance of the S&P 500 futures. The sample period is from July 1982 to September 2021, and the portfolio formation starts at January 2, 1990. The shaded areas denote the NBER recessions.

Table B1: **Alternative Market Proxies**

This table presents the regression of the daily stock market return (in excess of the risk-free rate) on the lagged signal extracted from the stocks owned by active mutual funds for alternative market proxies:

$$r^e_{m,t+1} = a_0 + a_1 s_t + a_2 r^e_{m,t} + \epsilon_{t+1},$$

where $r^e_{m,t+1}$ is the excess return of one of the five market proxies: S&P 500 Futures (FUT), S&P 500 E-mini Futures (Emini), S&P 500 ETF (SPDR), S&P 500 Total Return Index (Index), and CRSP Value-Weight Market Portfolio (VW). $s_t$ is the lagged difference between the equal-weighted average returns of the high-ownership and low-ownership stocks held by active mutual funds. $\{s_t\}$ is extracted from the universe of the all-but-micro-cap stocks, defined as the stocks with market cap above the 20th percentile of NYSE stocks. Newey and West (1987) $t$-statistics are reported in brackets. Coefficients that are significant at the 5% confidence level are in bold. The sample period is from February 1993 to September 2021.

| | $r^e_{FUT,t+1}$ | $r^e_{EMini,t+1}$ | $r^e_{SPDR,t+1}$ | $r^e_{Index,t+1}$ | $r^e_{VW,t+1}$ |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $s_t$ | **0.185** | **0.215** | **0.178** | **0.193** | **0.209** |
| | [3.82] | [3.56] | [3.71] | [3.93] | [4.20] |
| $r^e_{m,t}$ | **-0.102** | **-0.120** | **-0.109** | **-0.119** | **-0.098** |
| | [-3.80] | [-3.49] | [-3.76] | [-3.60] | [-2.91] |
| $R^2$ (%) | 7219 | 5777 | 7219 | 7219 | 7219 |
| N | 1.135 | 1.534 | 1.240 | 1.466 | 1.176 |

Table B2: **Control Other Predictors**

This table extends the findings in Table 1 by controlling for a variety of market return predictors that have already been proposed in the return predictability literature:

$$r^e_{m,t+1} = a_0 + a_1 s_t + a_2 control_t + \epsilon_{t+1},$$

where $r^e_{m,t+1}$ is the S&P 500 futures return on date $t+1$; $s_t$ is the difference between the equal-weighted average returns of the high-active-ownership and the low-active-ownership stocks on date $t$; and $control_t$ represents one of the following control variables available on date $t$: the aggregate turnover (TO) (Campbell et al., 1993); Variance Risk Premium (VRP) (Bollerslev et al., 2009); and Dividend Yield (DP), Earnings Yield (EP), Book-to-Market (BM), Inflation (INFL), Term Spread (TMS), Default Yield Spread (DFY), Net Equity Expansion (NTIS), all collected by Welch and Goyal (2008). The construction of the aggregate turnover follows Campbell et al. (1993). The data of VRP are downloaded from Prof. Hao Zhou's website and the data for the other control variables are downloaded from Prof. Amit Goyal's website. Newey and West (1987) $t$-statistics are reported in brackets. Coefficients that that are significant at the 5% confidence level are in bold. The sample period is from July 1982 to December 2020.

| | $r^e_{m,t+1}$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| $s_t$ | **0.219** | **0.202** | **0.220** | **0.220** | **0.220** | **0.219** | **0.219** | **0.219** | **0.219** | **0.202** |
| | [4.78] | [4.14] | [4.82] | [4.80] | [4.80] | [4.79] | [4.77] | [4.78] | [4.78] | [4.16] |
| $TO_t$ | 0.000 | | | | | | | | | -0.000 |
| | [0.29] | | | | | | | | | [-1.48] |
| $VRP_t$ | | 0.000 | | | | | | | | 0.000 |
| | | [0.14] | | | | | | | | [0.14] |
| $DP_t$ | | | **0.001** | | | | | | | 0.001 |
| | | | [2.43] | | | | | | | [1.82] |
| $EP_t$ | | | | 0.001 | | | | | | 0.000 |
| | | | | [1.40] | | | | | | [0.33] |
| $BM_t$ | | | | | **0.001** | | | | | -0.001 |
| | | | | | [2.05] | | | | | [-0.28] |
| $INFL_t$ | | | | | | -0.034 | | | | -0.037 |
| | | | | | | [-0.86] | | | | [-0.90] |
| $TMS_t$ | | | | | | | -0.003 | | | -0.025 |
| | | | | | | | [-0.33] | | | [-1.64] |
| $DFY_t$ | | | | | | | | 0.024 | | -0.027 |
| | | | | | | | | [0.59] | | [-0.33] |
| $NTIS_t$ | | | | | | | | | 0.000 | 0.015 |
| | | | | | | | | | [0.05] | [1.38] |
| $r^e_{m,t}$ | **-0.096** | **-0.101** | **-0.096** | **-0.096** | **-0.096** | **-0.096** | **-0.096** | **-0.096** | **-0.096** | **-0.102** |
| | [-4.11] | [-3.78] | [-4.12] | [-4.12] | [-4.12] | [-4.11] | [-4.10] | [-4.10] | [-4.11] | [-3.84] |
| N | 9708 | 7789 | 9708 | 9708 | 9708 | 9708 | 9708 | 9708 | 9708 | 7789 |
| $R^2$ (%) | 0.968 | 1.112 | 1.018 | 0.994 | 1.008 | 0.975 | 0.968 | 0.974 | 0.967 | 1.123 |

Table B3: **Signals Extracted From the 2-Month Lagged Active Ownership**

The SEC requires mutual funds to disclose their quarterly holdings no later than 60 days after the report date. To show the robustness of our main results, this table reproduces the results in Table 1 by using the 2-month lagged active mutual fund ownership to extract the signal for the S&P 500 futures returns.

| | $r^e_{m,t+1}$ | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $s_t$ | **0.108** | **0.191** | | | |
| | [3.46] | [4.50] | | | |
| $Sign(s_t)$ | | | **0.055** | | |
| | | | [4.13] | | |
| $s_t^+$ | | | | **0.264** | |
| | | | | [4.08] | |
| $s_t^-$ | | | | | **0.222** |
| | | | | | [3.29] |
| $r^e_{m,t}$ | | **-0.089** | **-0.075** | **-0.081** | **-0.078** |
| | | [-3.77] | [-3.40] | [-3.53] | [-3.46] |
| N | 9853 | 9853 | 9853 | 9853 | 9853 |
| $R^2$ (%) | 0.153 | 0.831 | 0.571 | 0.748 | 0.598 |
| $R^2_{OOS}$ (%) | $0.158^{***}$ | $0.829^{***}$ | $0.580^{***}$ | $0.730^{***}$ | $0.450^{***}$ |

Table B4: **Performance of the Daily Market-Timing Strategy: Signals Extracted From the 2-Month Lagged Active Ownership**

This table evaluates the performance of our daily out-of-sample market-timing strategy for trading the S&P 500 futures. The SEC requires mutual funds to disclose their quarterly holdings no later than 60 days after the report date. To show the robustness of our main results, this table reproduces the results in Table 3 by using the 2-month lagged active mutual fund ownership to extract the signal for the S&P 500 futures returns. Following Campbell and Thompson (2008) and Gao et al. (2018), our market-timing strategy is based on the optimal portfolio for a mean-variance investor with a risk aversion coefficient of 5, using our active-ownership signal. Specifically, our strategy adjusts the weight on the S&P 500 futures using the following formula:

$$w_t = \frac{\widehat{E}(r^e_{m,t+1})}{5 \times \widehat{Var}(r^e_{m,t+1})},$$

where $\widehat{E}(r^e_{m,t+1})$ is the out-of-sample expected return of the S&P 500 futures estimated from the predictive regression in Table 1 using the data from July 1982 to the date of portfolio formation and $\widehat{Var}(r^e_{m,t+1})$ is the out-of-sample variance estimated with a rolling window of 252 trading days. The weight is bounded between -0.5 and 1.5. The sample period is from July 1982 to September 2021, and portfolio formation starts in January 1990. Panel A presents the key statistics (i.e., $E(r^e_t)$ and Sharpe Ratio) and the certainty equivalent return (CER) gain of the trading strategy. The CER gain is the difference between the CER for an investor who uses the predictive regression forecast of the S&P 500 futures return and the CER for an investor who uses the historical average forecast, and it can be interpreted as the management fee per annum that the investor is willing to pay so as to be indifferent between investing in the market-timing strategy with the active-ownership signal versus an alternative market-timing strategy which estimates the out-of-sample equity premium with the in-sample average. Panel B evaluates the performance of the market-timing strategy against various benchmarks, including: CAPM, Carhart 4 factors (Carhart, 1997), Fama-French 5 factors (Fama and French, 2015), Q5 (Hou et al., 2019), and Daniel-Hirshleifer-Sun behavioral factors (Daniel et al., 2020). White (1980) Heteroskedasticity-robust $t$-statistics are reported in brackets. Coefficients that are significant at the 5% confidence level are in bold. "IR" denotes the annualized information ratios relative to the benchmarks.

Panel A. Performance of the Out-of-Sample Market-Timing Strategy

| $E(r_t^e)$ (%) | Std Dev (%) | SR | Skewness | Kurtosis | CER (%) |
|---|---|---|---|---|---|
| 14.14 | 15.92 | 0.89 | 0.93 | 44.95 | 4.66 |

Panel B. Regression of the Strategy's Excess Return on Alternative Factors

| | CAPM | Carhart | FF5 | Q5 | DHS |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $\alpha$ (%) | **11.18** | **10.74** | **10.69** | **10.78** | **10.66** |
| | [4.21] | [4.00] | [4.00] | [4.02] | [3.90] |
| $\beta_{mkt}$ | **0.32** | **0.33** | **0.34** | **0.34** | **0.30** |
| | [8.98] | [9.06] | [8.65] | [8.68] | [7.81] |
| $\beta_{smb}$ | | **-0.12** | **-0.10** | | |
| | | [-3.54] | [-3.08] | | |
| $\beta_{hml}$ | | 0.05 | 0.00 | | |
| | | [1.24] | [0.12] | | |
| $\beta_{umd}$ | | **0.07** | | | |
| | | [2.99] | | | |
| $\beta_{rmw}$ | | | 0.08 | | |
| | | | [1.79] | | |
| $\beta_{cma}$ | | | 0.06 | | |
| | | | [0.99] | | |
| $\beta_{r\_me}$ | | | | **-0.11** | |
| | | | | [-3.05] | |
| $\beta_{r\_ia}$ | | | | 0.05 | |
| | | | | [0.92] | |
| $\beta_{r\_roe}$ | | | | **0.14** | |
| | | | | [3.14] | |
| $\beta_{r\_eg}$ | | | | -0.04 | |
| | | | | [-0.87] | |
| $\beta_{mgmt}$ | | | | | 0.02 |
| | | | | | [0.38] |
| $\beta_{perf}$ | | | | | 0.05 |
| | | | | | [1.75] |
| IR | 0.75 | 0.73 | 0.72 | 0.73 | 0.75 |

Table B5: **Robustness: Alternative Financial Institutions, 1999 - 2020**

This table documents the predictability of the daily S&P 500 futures return by the lagged signals extracted from returns of stocks held by alternative financial institutions:

$$r^e_{m,t+1} = a_0 + a_1 s^{institution}_t + a_3 r^e_{m,t} + \epsilon_{t+1},$$

where $r^e_{m,t+1}$ is the S&P 500 futures return on date $t+1$ and $s^{institution}_t$ is the difference between the equal-weighted average returns of the stocks with high and low ownership held by a specific type of financial institution on date $t$. The signals are extracted from the universe of all-but-micro-cap stocks, defined as the stocks with market capitalization above the 20th percentile of NYSE stocks. Mutual fund and 13F institutional holdings are from the Thomson Reuters holdings data. Hedge fund holdings are from the FactSet Global Ownership data. The institution classification follows Koijen and Yogo (2019). Newey and West (1987) $t$-statistics are reported in brackets. Coefficients that are significant at the 5% confidence level are in bold. The sample period is from April 1999 to June 2020 when the holdings data for all institutions are available.

| | $r^e_{m,t+1}$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $s^{institution}_t$ | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Active Mutual Fund | **0.217** | | | | | | | **0.275** | **0.238** |
| | [3.64] | | | | | | | [3.54] | [3.04] |
| Passive Fund and ETF | | 0.056 | | | | | | 0.131 | 0.146 |
| | | [1.18] | | | | | | [1.39] | [1.55] |
| Investment Advisor | | | 0.115 | | | | | 0.014 | -0.161 |
| | | | [1.48] | | | | | [0.17] | [-1.92] |
| Pension Fund | | | | 0.047 | | | | 0.153 | 0.173 |
| | | | | [0.81] | | | | [1.49] | [1.66] |
| Bank | | | | | -0.007 | | | -0.081 | -0.002 |
| | | | | | [-0.16] | | | [-0.97] | [-0.02] |
| Insurance Company | | | | | | 0.001 | | **-0.319** | **-0.311** |
| | | | | | | [0.01] | | [-2.23] | [-2.18] |
| Hedge Fund | | | | | | | **0.153** | | **0.227** |
| | | | | | | | [2.18] | | [2.25] |
| $r^e_{m,t}$ | **-0.102** | **-0.084** | **-0.083** | **-0.082** | **-0.083** | **-0.083** | **-0.099** | **-0.112** | **-0.128** |
| | [-3.22] | [-2.79] | [-2.76] | [-2.73] | [-2.71] | [-2.78] | [-2.94] | [-3.78] | [-3.87] |
| N | 5347 | 5347 | 5347 | 5347 | 5347 | 5347 | 5347 | 5347 | 5347 |
| $R^2$ (%) | 1.247 | 0.726 | 0.792 | 0.694 | 0.653 | 0.652 | 0.989 | 1.652 | 1.844 |