

Tomorrow Is Another Day: Stocks Active Mutual Funds Overweight Predict the Next-Day Market *

Shuaiyu Chen[†] Yixin Chen[‡] Randolph B. Cohen[§]

First Draft: August 16, 2021
This Version: November 5, 2021

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 does 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 over 0.9. 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. 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 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.¹

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

¹Even if active mutual funds might play a zero-sum game for trading profits, it 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.

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 exhibit predictive power; whereas the holdings of 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 and higher trading volume. 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 the case, in particular for the industries with high active-mutual-fund participation. Lastly, at the individual stock level, 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 docu-

ment 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, Xue, and Zhang \(2020\)](#) 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, Cohen, and Wang \(2020\)](#) showing abnormal market performance on certain pre-determined dates; [Bollerslev, Tauchen, and Zhou \(2009\)](#), [Ross \(2015\)](#), [Martin \(2017\)](#) relating market return predictability to the pricing of derivative contracts; [Campbell, Grossman, and Wang \(1993\)](#), [Kelly and Pruitt \(2013\)](#), [Huang, Jiang, Tu, and Zhou \(2015\)](#), [Rapach, Ringgenberg, and Zhou \(2016\)](#), [Jiang, Lee, Martin, and Zhou \(2019\)](#), [Engelberg, McLean, Pontiff, and Ringgenberg \(2019\)](#), [Dong, Li, Rapach, and Zhou \(2021\)](#) 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, Yao, and Yu \(2007\)](#); [Kacperczyk, Nieuwerburgh, and Veldkamp \(2014\)](#)). 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.

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 discusses and distinguishes the competing hypotheses for our empirical findings; Section 6 concludes.

2. Theoretical Framework

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

2.1. Model Setup

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

2.1.1. Securities

The market is segmented with two groups of risky securities. There is a group of N stocks held by active managers (the high-AO stocks); and a group of N stocks held by uninformed investors (the low-AO stocks). Each stock pays out a stream of dividend:

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

where $D_{i,t}^{Hi}(D_{i,t}^{Lo})$ 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 $\epsilon_{a,t+1} \sim \mathbb{N}(0, \sigma_a^2)$ is the aggregate component in the dividend process; and $d_{i,t+1}^{Hi} = \phi d_{i,t}^{Hi} + \epsilon_{i,t+1}^{Hi}$ ($d_{i,t+1}^{Lo} = \phi d_{i,t}^{Lo} + \epsilon_{i,t+1}^{Lo}$) with $\epsilon_{i,t+1}^{Hi} \sim \mathbb{N}(0, \sigma_i^2)$ ($\epsilon_{i,t+1}^{Lo} \sim \mathbb{N}(0, \sigma_i^2)$) is the stock-specific component. The random cash flow shocks $\{\epsilon_{a,t}\}_{t=0}^{\infty}, \{\epsilon_{i,t}^{Hi}\}_{t=0}^{\infty}, \{\epsilon_{i,t}^{Lo}\}_{t=0}^{\infty}$ are i.i.d normal.

All risky securities are of unit supply.

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

2.1.2. 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 have the same preference. Each investor maximizes over

$$J(W_t) = \max_{\{C_s, X_s\}} \sum_{s=t}^{\infty} \beta^{s-t} \mathbb{E}(U(C_s) | \mathcal{F}_t), \quad (2)$$

where

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

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 ; X_t represents the investor's portfolio choice; $\beta \in (0, 1)$ is the subject discount rate; $\alpha^I(\alpha^U)$ 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,

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

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.

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

2.2. Equilibrium

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

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

the price of a low-AO stock is

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

The value function of an informed asset manager take the form

$$J^i(W_t) = -\exp(-a^I W_t - b^I), \tag{8}$$

and the value function of the uninformed investor is

$$J^U(W_t) = -\exp(-a^U W_t - b^U) \tag{9}$$

for some constants: $A^{Hi}, B^{Hi}, C^{Hi}, A^{Lo}, B^{Lo}, C^{Lo}, D^{Lo}, a^I, b^I, a^U$, and b^U .

Proof. See Appendix A. □

Corollary 1. *The average price of the high-AO stock is*

$$\begin{aligned} P_t^{Hi} = & \frac{1}{R-1} \left(\mu - A^{Hi} \right) + \frac{\phi}{R-\phi} \left(d_{a,t} + \frac{\sum_i d_{i,t}^{Hi}}{N} \right) \\ & + B^{Hi} \left(\epsilon_{a,t+1} + \frac{\sum_i \epsilon_{i,t+1}^{Hi}}{N} \right) + C^{Hi} \frac{\sum_i u_{i,t}^{Hi}}{N}. \end{aligned} \quad (10)$$

The average price of the low-AO stock is

$$\begin{aligned} P_t^{Lo} = & \frac{1}{R-1} \left(\mu - A^{Lo} \right) + \frac{\phi}{R-\phi} \left(d_{a,t} + \frac{\sum_i d_{i,t}^{Lo}}{N} \right) \\ & + C^{Lo} \left[B^{Hi} \left(\frac{\sum_i \epsilon_{i,t+1}^{Hi}}{N} + \epsilon_{a,t+1} \right) + C^{Hi} \frac{\sum_i u_{i,t}^{Hi}}{N} \right] + D^{Lo} \frac{\sum_i u_{i,t}^{Lo}}{N}. \end{aligned} \quad (11)$$

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

$$\begin{aligned} R_t^{Hi} \equiv & P_t^{Hi} + D_t^{Hi} - P_{t-1}^{Hi} \\ = & \mu + \frac{\phi}{R-\phi} \left[\left((R-1) d_{a,t-1} + \frac{R}{\phi} \epsilon_{a,t} \right) + \left((R-1) \frac{\sum_i d_{i,t-1}^{Hi}}{N} + \frac{R}{\phi} \frac{\sum_i \epsilon_{i,t}^{Hi}}{N} \right) \right] \\ & + B^{Hi} \left[\left(\epsilon_{a,t+1} + \frac{\sum_i \epsilon_{i,t+1}^{Hi}}{N} \right) - \left(\epsilon_{a,t} + \frac{\sum_i \epsilon_{i,t}^{Hi}}{N} \right) \right] \\ & + C^{Hi} \left(\frac{\sum_i u_{i,t}^{Hi}}{N} - \frac{\sum_i u_{i,t-1}^{Hi}}{N} \right). \end{aligned} \quad (12)$$

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

$$\begin{aligned} R_t^{Lo} \equiv & P_t^{Lo} + D_t^{Lo} - P_{t-1}^{Lo} \\ = & \mu + \frac{\phi}{R-\phi} \left[\left((R-1) d_{a,t-1} + \frac{R}{\phi} \epsilon_{a,t} \right) + \left((R-1) \frac{\sum_i d_{i,t-1}^{Lo}}{N} + \frac{R}{\phi} \frac{\sum_i \epsilon_{i,t}^{Lo}}{N} \right) \right] \\ & + C^{Lo} B^{Hi} \left[\left(\epsilon_{a,t+1} + \frac{\sum_i \epsilon_{i,t+1}^{Hi}}{N} \right) - \left(\epsilon_{a,t} + \frac{\sum_i \epsilon_{i,t}^{Hi}}{N} \right) \right] \\ & + C^{Lo} C^{Hi} \left(\frac{\sum_i u_{i,t}^{Hi}}{N} - \frac{\sum_i u_{i,t-1}^{Hi}}{N} \right) + D^{Lo} \left(\frac{\sum_i u_{i,t}^{Lo}}{N} - \frac{\sum_i u_{i,t-1}^{Lo}}{N} \right). \end{aligned} \quad (13)$$

Proof. Immediate from Proposition 1. $\square$

Proposition 2. *The average price of the high-AO stocks are more informed about the next-period aggregate shock, $\epsilon_{a,t+1}$, than the average price of the low-AO stocks, i.e. $\text{Var}(\epsilon_{a,t+1}|P_t^{Hi}) < \text{Var}(\epsilon_{a,t+1}|P_t^{Lo})$.*

Proof. Immediate from Corollary 1 and the fact that $D^{Lo} \neq 0$ and $\text{Var}\left(\frac{\sum_i u_{i,t}^{Lo}}{N}\right) > 0$. $\square$

Proposition 3. *Define s_t as the relative performance between the high-AO and low-AO stocks, i.e.*

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

Under the parameter specification where $\text{Var}(R_t^{Hi}) = \text{Var}(R_t^{Lo})$ ², s_t positively predicts the next-period aggregate shock, $\epsilon_{a,t+1}$, and thus the next-period market return.

Proof. According to Corollary 1,

$$\begin{aligned} s_t &\equiv R_t^{Hi} - R_t^{Lo} \\ &= \frac{\phi}{R - \phi} \left((R - 1) \frac{\sum_i d_{i,t-1}^{Hi} - \sum_i d_{i,t-1}^{Lo}}{N} + \frac{R}{\phi} \frac{\sum_i \epsilon_{i,t}^{Hi} - \sum_i \epsilon_{i,t}^{Lo}}{N} \right) \\ &\quad + \left(1 - C^{Lo}\right) B^{Hi} \left[\left(\epsilon_{a,t+1} + \frac{\sum_i \epsilon_{i,t+1}^{Hi}}{N} \right) - \left(\epsilon_{a,t} + \frac{\sum_i \epsilon_{i,t}^{Hi}}{N} \right) \right] \\ &\quad + \left(1 - C^{Lo}\right) C^{Hi} \left(\frac{\sum_i u_{i,t}^{Hi}}{N} - \frac{\sum_i u_{i,t-1}^{Hi}}{N} \right) + D^{Lo} \left(\frac{\sum_i u_{i,t}^{Lo}}{N} - \frac{\sum_i u_{i,t-1}^{Lo}}{N} \right) \\ &\approx \left(1 - C^{Lo}\right) B^{Hi} \epsilon_{a,t+1} + \xi_t \quad \text{when } N \text{ is sufficiently large,} \end{aligned}$$

where $\xi_t = (1 - C^{Lo}) [C^{Hi} (u_t^{Hi} - u_{t-1}^{Hi}) - B^{Hi} \epsilon_{a,t}] + D^{Lo} (u_t^{Lo} - u_{t-1}^{Lo})$.

Also, $C^{Lo} < 1$ when $\text{Var}(R_t^{Hi}) = \text{Var}(R_t^{Lo})$. $\square$

2.3. Remarks

Despite the lengthy algebra, the intuition of the model is straightforward.

In the model, each informed active manager studies a particular stock. The signal received by the manager is about a specific stock only and is not about the

²In the data, the daily volatility of the high-AO portfolio is 1.31%, which is close to the low-AO portfolio volatility of 1.15%.

aggregate market directly. Each individual manager then takes advantage of the stock-level information that is available to her, and is only active at trading and incorporating such 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; and 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.

And finally, subtracting the low-AO portfolio return R_t^{Lo} from the high-AO portfolio return R_t^{Hi} removes common components such as $d_{a,t-1}$ and $\epsilon_{a,t}$, so that the relative performance between the two is a strong signal that predicts the next-period market.

3. Data

Data in this paper are from several sources. The US and global equity data 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: 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 exchange rates from Compustat.

For the proxies of the aggregate equity markets, we collect the daily price data of the most liquid and traded futures for the 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 the market returns (see., [Moskowitz, Ooi, and Pedersen, 2012](#); [Koijen, Moskowitz, Pedersen, and Vrugt, 2018](#)).

The data of US equity mutual funds are from 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 (ETFs). The active funds are identified based on the screening procedure used in [Kacperczyk, Sialm, and Zheng \(2008\)](#) and [Cremers and Pareek \(2016\)](#).³ To identify index funds and ETFs, we first rely on the indicator of fund types in CRSP, and then screen by fund names following the procedure proposed by [Appel, Gormley, and Keim \(2016\)](#).⁴ To mitigate the incubation bias raised 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 the 2006 dollar ([Fama and French, 2010](#)). [Zhu \(2020\)](#) documents 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 Ser-

³We require 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 is not focusing 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.

⁴We 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 \(2020\)](#) shows 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.

vices (WRDS) SEC Analytics. Ben-David, Franzoni, Moussawi, and Sedunov (2021) point out several data issues in the Thomson Reuters Database and conduct an assessment of the potential biases caused by these issues. Following their suggestion, we use the Thomson Reuters data before June 2013 and the SEC 13F filings data parsed by WRDS SEC Analytics afterwards.⁶ We identify equity holdings of hedge funds based on the institution classification in the FactSet Global Ownership Database (Ben-David, Franzoni, and Moussawi, 2012).

Our data of global mutual fund holdings are from the FactSet Global Ownership Database, which is widely used in the recent international studies on mutual funds (see, Cremers, Ferreira, Matos, and Starks, 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.

For the intraday analyses in our paper, we process the intraday transactions data from NYSE Trade and Quote (TAQ) database. We construct analyst coverage and media coverage based on the data from IEBS and RavenPack, respectively. For most of the exercises on the US equity market, the sample period is from January 1983 to December 2020 when 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.

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

4. Empirical Results

4.1. Main Findings

4.1.1. Signal Construction

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

1. We exclude micro-cap stocks with market cap below the 20th percentile of NYSE stocks according to the definition in [Fama and French \(2008\)](#).
2. Among the remaining all-but-micro-cap stocks, we sort them into five groups by AO 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 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 the following considerations. We exclude the micro-cap stocks because, for a tiny stock, the behavior of its share price tends to be erratic ([Fama and French, 2008](#); [Hou et al., 2020](#)), and we attempt to remove the influence of such micro-cap stocks on our signal. We take the equal-weighted average of stock returns because the distribution of individual stock market capitalization is highly right-skewed in the cross section, so that 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, we take the difference in the average

⁷Lagging the holdings by two months has little effect on our results. See Table [B2](#) and Table [B3](#) in Appendix.

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

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

returns between the high-AO and low-AO groups to remove any market-wide common components in stock returns that are not predictive of the subsequent market return but could potentially contaminate our signal.¹⁰

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

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 0.14, with a Newey and West (1987) t -statistic of 4.12, and the predictive regression has an R-square of 0.26%. Column (2) shows that controlling for the lagged market return further strengthens the result, where the coefficient on the signal rises substantially to 0.22 (with a t -statistic of 4.72), and the daily R-square reaches an impressive level of 1%. 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.

[Insert Table 1 near here]

To ensure our signal is fully tradable and is not driven by market microstructure or liquidity issues, we adopt the S&P 500 futures as our proxy of the stock market for most of our empirical exercises. Some questions around this empirical choice are whether our results are robust and how they would change if alternative financial instruments of the S&P 500 index are used. Table B1 in Appendix shows that our results are virtually unchanged when we consider alternative market proxies such as the S&P 500 E-mini futures, the S&P 500 ETF, or the CRSP value-weighted market.

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

4.1.3. *International Evidence*

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

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

[Insert Table 2 near here]

Therefore, our key finding is not isolated to the US, but is rather a prevalent phenomenon around the world. For the rest of the paper, 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*

Our finding of the strong daily S&P 500 futures return predictability is a striking result because the 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 strategy based on our signal and evaluate its performance using different factor models.

Following Campbell and Thompson (2008) and Gao, Han, Li, and Zhou (2018),

we construct the optimal portfolio for a mean-variance investor with a risk aversion coefficient of five using our active ownership signal. The weight on the S&P 500 futures is bounded between -0.5 and 1.5. The out-of-sample equity premium is estimated with in-sample data since the beginning of the sample until the date of portfolio formation; and the out-of-sample volatility is estimated with a rolling window of 252 trading days. The first seven years of the data is 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 presents that, during the sample over the last 30 years, the strategy realizes a premium of 15% per year, with an impressive Sharpe ratio of 0.91. The utility gain to the mean-variance investor is equivalent to a 6.71% 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 exposures to popular risk factors. The information ratio of the strategy ranges from 0.75 to 0.79, 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 and financial crises.

Therefore, our finding of the strong predictability of the S&P 500 futures return by the 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 finding of the 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 is inconsistent with these channels. Then we present the additional findings that can strongly support 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 performances 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, either over the subsequent five days or 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 components¹¹; Panel B of the table shows that our signal's predictability mainly stems from its intraday component on the intraday component of the next-day market return.

The lack of the subsequent reversal in the price of the market and the intraday timing of the predictive power of the signal thus can rule out the temporary price pressure channel. Indeed, such a channel predicts the reversal in market price; and since the buying of the mutual fund managers occurs overnight, the channel would also predict the overnight market return predictability.

¹¹Following Bogousslavsky (2021), we take the price at 9:45am as the open price to mitigate potential microstructure issues; for individual stock prices which we use to generate the signal, open price is defined as the midquote at 9:45am.

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 stocks with high active-mutual-fund ownership adjust faster and thus can 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 it is the stocks with these alternative confounding firm characteristics producing the signal that predicts the market.¹²

[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, high-active-ownership stocks 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 high active ownership are smaller than the stocks in the middle. This is an intuitive finding because, on the one hand, mutual funds are reluctant to hold illiquid stocks which tend to be small; on the other hand, their allocation to small stocks tend 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 with an alternative firm characteristic. The table shows that only the signal generated by active-mutual-fund ownership demonstrates significant predictive power for the market. Therefore, we conclude that the predictability

¹²For the empirical relations between active mutual fund ownership and firm characteristics, see, [Falkenstein \(1996\)](#), [Bennett, Sias, and Starks \(2003\)](#), [Massa, Phalippou, et al. \(2004\)](#), [Cao, Simin, and Wang \(2013\)](#), [Solomon, Soltes, and Sosyura \(2014\)](#), [Fang, Peress, and Zheng \(2014\)](#), etc.

of our signal is not due to the correlation between the active ownership measure and other confounding firm characteristics.

4.2.3. *Predictability by Alternative Financial Institutions*

In our preferred explanation, the daily market return predictability by 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 the active mutual fund sector 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 difference across the seven major types of institutional investors: active mutual fund, passive mutual 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 fund, investment advisor (which include asset management company) and hedge fund show predictive power for the market return. Intuitively, the stock ownership by pension fund, bank, and insurance company does not help to predict the market return at a daily frequency as these institutions do not engage in taking market-timing bets on a daily basis.

Importantly, the signal associated with the ownership by passive fund and ETF does not predict the market return either. Such a finding suggests that the predictive power of our signal indeed 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. It is also striking that there is a drastic difference in the market return predictive power between stocks' 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 funds. But this need not be the case; if the market reacts quickly to the trades of active managers, they can be the channel by which information makes its way into the market while receiving only very modest performance benefits for their service (Grossman and Stiglitz, 1980; Kyle, 1985).

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 that 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 8 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 Composite Sentiment Score (CSS) of business-related news articles across all firms (dubbed as the "aggregate stock news sentiment"). Panel A of Table 8 shows that, consistent with our model, our AO signal is indeed able to predict the next-day aggregate stock news sentiment. Moreover, in Panel B, we find that the predicted aggregate stock news sentiment is significantly positively correlated with the next-day market return. Therefore, these findings strongly support our information channel by showing that, at least, part of the AO signal's predictive power of the market return is derived from its ability to predict the next-day aggregated stock-level news sentiment.

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

[Insert Table 9 near here]

Another implication of our model of active fund manager acumen is that the effect should be stronger if we isolate the managers who have shown 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 Carhart (1997) benchmark in the prior 24-month rolling window. Out-of-sample signals are produced from the ownership by these two groups of active funds, separately. Panel A in Table 9 shows that, consistent with our conjecture, the signal constructed from the funds with high historical information ratio demonstrates much stronger predictive power than the signal produced from the low-information-ratio-fund ownership. The predictive power of the high-information-ratio signal is comparable to the full version of the signal when all active mutual funds are included.

In the same spirit, since the predictability of our signal is at daily frequency, we should also expect its effectiveness to come mostly from the funds that make high-frequency bets on stocks. Also, previous studies have documented that high-turnover funds tend to outperform low-turnover funds because of their superior ability of exploiting time-varying investment opportunities (Pástor, Stambaugh, and Taylor, 2017). Motivated by this conjecture, we partition the active mutual funds into two groups by their prior-year turnover ratio, and construct the signals based on the high-turnover AO and low-turnover AO, separately. Panel B in Table 9 verifies this conjecture by showing that the predictability of the signal indeed derives from the high-turnover funds within the active mutual fund sector.

4.3.3. *Daily Industry Return Predictability*

Kacperczyk, Sialm, and Zheng (2005) document that active mutual fund managers may possess private information about certain industries; and consequently,

tilt their portfolio weights towards these industries. So in addition to our main empirical findings regarding the daily market return predictability, we also extend the same logic and study the predictability of industry-specific returns. To deviate from the market return predictability exercise, we measure the daily 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 the industry-specific signals by only including stocks within the specific industries when constructing the signals.

[Insert Table 10 near here]

Table 10 presents the predictive regressions with the industry-specific returns and signals. Consistent with our intuition, we find strong return predictability for industries with high active-mutual-fund ownership, such as finance or business services; but no predictability is found for industries with less mutual fund participation, such as telecommunication 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 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 individual stocks and study the lead-lag relations among their returns in the same spirit as Lo and MacKinlay (1990) and Hameed, Lof, and Suominen (2017). If the prices of the high-AO stocks are indeed more efficient than the other stocks, then we should expect cross-predictability from the returns of the high-AO stocks to those of the low-AO stocks, but not vice versa. To ensure such a lead-lag relation is tradable, following Bogousslavsky (2021), we predict the stock returns from 9:45am to market close with the close-to-close returns on the previous trading day.

[Insert Table 11 near here]

Table 11 confirms the conjecture and shows the strong return cross-predictability among the stocks sorted by active ownership. The daily returns of the high-AO stocks positively predict the returns of the other stocks. 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 unhelpful market-wide common components and distill a signal that has the strongest predictive power. The high-AO portfolio also demonstrates strong momentum at a daily frequency. Such a finding is consistent with the [Lo and MacKinlay \(1990\)](#) channel that the lead-lag relation within the high-AO group can generate the momentum effect for the group as a whole.

5. Discussion 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 market or industry return. We discuss below 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, the 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 manager possesses a sufficiently accurate market-timing signal to profit net of trading costs; that only the combined wisdom of all, or at least many, managers suffices to effectively forecast market performance. A likely explanation is that active managers possess information only about the individual stocks they care about, but that in aggregate these shards of information combine to forecast the market. Of course it's also possible that one form active-manager information takes is indeed market-wide; perhaps in-

formation about future Fed action or broad economic trends gets to, or is processed by, active managers sooner than other market participants.

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. And then the owners of the other stocks could 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 hedge funds but not from those primarily held by other types of financial institutions such as banks, insurance companies, or pension funds. Such an observation 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 that we do not expect high-frequency predictability of the market in the returns of the stocks held by them. Moreover, within the active mutual fund industry, we identify those funds with better performance and high trading volume as the source of the predictive signal. These observations further strengthen the support for our channel, because consistent with the channel, we indeed expect the more skilled funds to incorporate their private information better, and since our signal is predictive at 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 or 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 the 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, [Edelen and Warner \(2001\)](#) show, using proprietary data, that the daily aggregate mutual fund flows correlate with the market returns concurrently or with a lag but do not predict the 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 thus temporarily push up (down) the security prices. However, this temporary price pressure channel is inconsistent with our intraday analysis, where we find that our signal mostly predicts the open-to-close market return the next day rather than the close-to-open return. The flow pressure to the funds should induce them to trade overnight or at the next-day market open so as to generate the predictability of overnight market return instead. In addition, although the temporary price pressure could generate the momentum effect for the high-AO stocks, it does not cause the lead-lag relation between the high- and low-AO stocks. Finally, the return gap between high- and low-AO stocks predicts the next-day aggregate stock-level news sentiment, a measure that is correlated with market returns but not directly affected by trading. This result is entirely consistent with the idea that when high-AO stocks outperform, stocks generally will experience good news the next day; moreover, the ability

of the AO signal to predict the next-day market sentiment is, we think, difficult to explain any other way.

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

Other questions of great interest concern the source and type of information active mutual fund managers are employing, and the mechanism by which that information becomes incorporated into prices. For example, to what extent is the information about the individual stocks the managers hold and study, and to what extent is it about the market as a whole (economic growth, Fed action, etc.)? Is the information primarily a result of active managers skillfully interpreting public announcements on their own, or is 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 so they must await future research.

6. 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 on 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, rather, 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 provide 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 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 than that fact implies. So in sum, our findings imply that the stock market might be less efficient, and the active mutual funds might be more informed than what common beliefs would suggest.

References

- Appel, I. R., Gormley, T. A., Keim, D. B., 2016. Passive investors, not passive owners. *Journal of Financial Economics* 121, 111–141.
- Ben-David, I., Franzoni, F., Moussawi, R., 2012. Hedge fund stock trading in the financial crisis of 2007–2009. *The Review of Financial Studies* 25, 1–54.
- Ben-David, I., Franzoni, F., Moussawi, R., Sedunov, J., 2021. The granular nature of large institutional investors. *Management Science* .
- Bennett, J. A., Sias, R. W., Starks, L. T., 2003. Greener pastures and the impact of dynamic institutional preferences. *The Review of Financial Studies* 16, 1203–1238.
- Bogousslavsky, V., 2021. The cross-section of intraday and overnight returns. *Journal of Financial Economics* .
- Bollen, N. P., Busse, J. A., 2001. On the timing ability of mutual fund managers. *The Journal of Finance* 56, 1075–1094.
- Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. *The Review of Financial Studies* 22, 4463–4492.
- Campbell, J. Y., Grossman, S. J., Wang, J., 1993. Trading volume and serial correlation in stock returns. *The Quarterly Journal of Economics* 108, 905–939.
- Campbell, J. Y., Thompson, S. B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies* 21, 1509–1531.
- Cao, C., Simin, T. T., Wang, Y., 2013. Do mutual fund managers time market liquidity? *Journal of Financial Markets* 16, 279–307.
- Carhart, M. M., 1997. On persistence in mutual fund performance. *The Journal of finance* 52, 57–82.
- Chen, Y., Cohen, R. B., Wang, Z. K., 2020. Famous firms, earnings clusters, and the stock market. *Earnings Clusters, and the Stock Market* (September 2, 2020) .
- Clark, T. E., West, K. D., 2007. Approximately normal tests for equal predictive accuracy in nested models. *Journal of econometrics* 138, 291–311.
- Cremers, M., Ferreira, M. A., Matos, P., Starks, L., 2016. Indexing and active fund management: International evidence. *Journal of Financial Economics* 120, 539–560.
- Cremers, M., Pareek, A., 2016. Patient capital outperformance: The investment skill of high active share managers who trade infrequently. *Journal of Financial Economics* 122, 288–306.
- Dong, X., Li, Y., Rapach, D., Zhou, G., 2021. Anomalies and the expected market return. *Baruch College Zicklin School of Business Research Paper* p. 02.

- Edelen, R. M., Warner, J. B., 2001. Aggregate price effects of institutional trading: a study of mutual fund flow and market returns. *Journal of financial Economics* 59, 195–220.
- Engelberg, J., McLean, R. D., Pontiff, J., Ringgenberg, M., 2019. Are cross-sectional predictors good market-level predictors? In: *American Finance Association Annual Meeting Paper*.
- Evans, R. B., 2010. Mutual fund incubation. *The Journal of Finance* 65, 1581–1611.
- Falkenstein, E. G., 1996. Preferences for stock characteristics as revealed by mutual fund portfolio holdings. *The Journal of Finance* 51, 111–135.
- Fama, E. F., French, K. R., 2008. Dissecting anomalies. *The Journal of Finance* 63, 1653–1678.
- Fama, E. F., French, K. R., 2010. Luck versus skill in the cross-section of mutual fund returns pp. 261–300.
- Fang, L. H., Peress, J., Zheng, L., 2014. Does media coverage of stocks affect mutual funds' trading and performance? *The Review of Financial Studies* 27, 3441–3466.
- Gao, L., Han, Y., Li, S. Z., Zhou, G., 2018. Market intraday momentum. *Journal of Financial Economics* 129, 394–414.
- Greenwood, R., Hanson, S. G., 2012. Share issuance and factor timing. *The Journal of Finance* 67, 761–798.
- Greenwood, R., Hanson, S. G., 2013. Issuer quality and corporate bond returns. *The Review of Financial Studies* 26, 1483–1525.
- Grossman, S. J., Stiglitz, J. E., 1980. On the impossibility of informationally efficient markets. *The American economic review* 70, 393–408.
- Hameed, A., Lof, M., Suominen, M., 2017. Slow trading and stock return predictability .
- Henriksson, R. D., Merton, R. C., 1981. On market timing and investment performance. ii. statistical procedures for evaluating forecasting skills. *Journal of business* pp. 513–533.
- Hou, K., Xue, C., Zhang, L., 2020. Replicating anomalies. *The Review of Financial Studies* 33, 2019–2133.
- Huang, D., Jiang, F., Tu, J., Zhou, G., 2015. Investor sentiment aligned: A powerful predictor of stock returns. *The Review of Financial Studies* 28, 791–837.
- Jiang, F., Lee, J., Martin, X., Zhou, G., 2019. Manager sentiment and stock returns. *Journal of Financial Economics* 132, 126–149.
- Jiang, G. J., Yao, T., Yu, T., 2007. Do mutual funds time the market? evidence from portfolio holdings. *Journal of Financial Economics* 86, 724–758.

- Kacperczyk, M., Nieuwerburgh, S. V., Veldkamp, L., 2014. Time-varying fund manager skill. *The Journal of Finance* 69, 1455–1484.
- Kacperczyk, M., Sialm, C., Zheng, L., 2005. On the industry concentration of actively managed equity mutual funds. *The Journal of Finance* 60, 1983–2011.
- Kacperczyk, M., Sialm, C., Zheng, L., 2008. Unobserved actions of mutual funds. *The Review of Financial Studies* 21, 2379–2416.
- Kelly, B., Pruitt, S., 2013. Market expectations in the cross-section of present values. *The Journal of Finance* 68, 1721–1756.
- Koijen, R. S., Moskowitz, T. J., Pedersen, L. H., Vrugt, E. B., 2018. Carry. *Journal of Financial Economics* 127, 197–225.
- Koijen, R. S., Yogo, M., 2019. A demand system approach to asset pricing. *Journal of Political Economy* 127, 1475–1515.
- Kyle, A. S., 1985. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society* pp. 1315–1335.
- Lakonishok, J., Smidt, S., 1988. Are seasonal anomalies real? a ninety-year perspective. *The review of financial studies* 1, 403–425.
- Lewellen, J., 2014. The cross section of expected stock returns. Forthcoming in *Critical Finance Review*, Tuck School of Business Working Paper .
- Lo, A. W., MacKinlay, A. C., 1990. When are contrarian profits due to stock market overreaction? *The review of financial studies* 3, 175–205.
- Lucca, D. O., Moench, E., 2015. The pre-fomc announcement drift. *The Journal of finance* 70, 329–371.
- Malkiel, B. G., 1995. Returns from investing in equity mutual funds 1971 to 1991. *The Journal of finance* 50, 549–572.
- Martin, I., 2017. What is the expected return on the market? *The Quarterly Journal of Economics* 132, 367–433.
- Massa, M., Phalippou, L., et al., 2004. Mutual funds and the market for liquidity. Tech. rep., CEPR Discussion Papers.
- McLean, R. D., Pontiff, J., 2016. Does academic research destroy stock return predictability? *The Journal of Finance* 71, 5–32.
- Moskowitz, T. J., Ooi, Y. H., Pedersen, L. H., 2012. Time series momentum. *Journal of financial economics* 104, 228–250.
- Newey, W. K., West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Pástor, L., Stambaugh, R. F., Taylor, L. A., 2017. Do funds make more when they trade more? *The Journal of Finance* 72, 1483–1528.

- Rapach, D. E., Ringgenberg, M. C., Zhou, G., 2016. Short interest and aggregate stock returns. *Journal of Financial Economics* 121, 46–65.
- Ross, S., 2015. The recovery theorem. *The Journal of Finance* 70, 615–648.
- Savor, P., Wilson, M., 2013. How much do investors care about macroeconomic risk? evidence from scheduled economic announcements. *Journal of Financial and Quantitative Analysis* 48, 343–375.
- Schumacher, D., 2018. Home bias abroad: Domestic industries and foreign portfolio choice. *The Review of Financial Studies* 31, 1654–1706.
- Solomon, D. H., Soltes, E., Sosyura, D., 2014. Winners in the spotlight: Media coverage of fund holdings as a driver of flows. *Journal of Financial Economics* 113, 53–72.
- Wang, J., 1993. A model of intertemporal asset prices under asymmetric information. *The Review of Economic Studies* 60, 249–282.
- White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: journal of the Econometric Society* pp. 817–838.
- Zhu, Q., 2020. The missing new funds. *Management Science* 66, 1193–1204.

Value of \$1 Invested in the Market-Timing Strategy

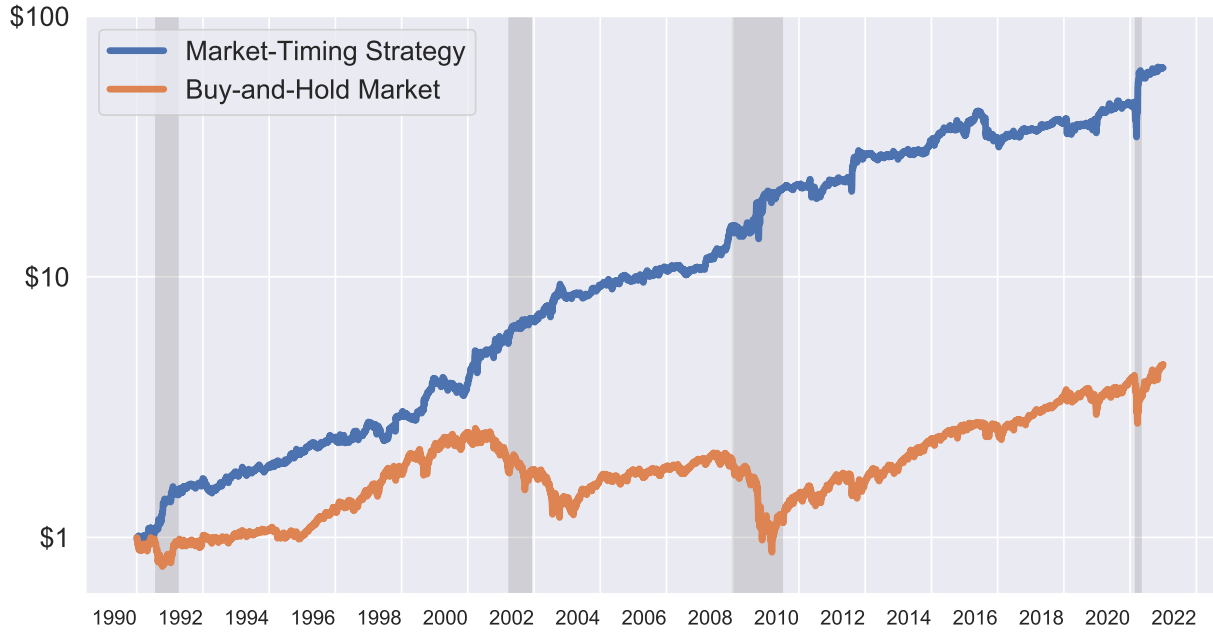


Fig. 1. Performance of the Daily Market-Timing Strategy

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. \(2018\)](#), we construct the optimal portfolio for a mean-variance investor with a risk aversion coefficient of five using our active-ownership signal. The weight on the S&P 500 futures is bounded between -0.5 and 1.5 . The out-of-sample equity premium is estimated with in-sample data since January 1983 until the date of portfolio formation; the out-of-sample volatility is estimated with a rolling window of 252 trading days. 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 January 1983 to December 2020, and portfolio formation starts from January 1990. The shaded areas denote the NBER recessions.

Table 1: Daily S&P 500 Futures Predictability

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

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

where $r_{m,t+1}^e$ 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 low-active-ownership stocks on date t ; $f(s_t)$ is a monotonic transformation of s_t , including: s_t itself, the sign of s_t , the positive component of s_t ($s_t^+ = \max(s_t, 0)$), and the negative component of s_t ($s_t^- = \min(s_t, 0)$). $\{s_t\}$ 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 with 50 lags are reported. Coefficients that are significant at the 5% confidence level are in bold. R_{OOS}^2 is the Campbell and Thompson (2008) out-of-sample R^2 statistic, whose statistical significance is based on the p -value for the Clark and West (2007) out-of-sample MSPE-adjusted statistic. “***” indicates significance at the 1% level. The sample period is from January 1983 to December 2020.

	$r_{m,t+1}^e$				
	(1)	(2)	(3)	(4)	(5)
s_t	0.135 [4.12]	0.224 [4.72]			
$Sign(s_t)$			0.070 [4.43]		
s_t^+				0.300 [4.11]	
s_t^-					0.271 [4.12]
$r_{m,t}^e$		-0.094 [-3.83]	-0.078 [-3.40]	-0.084 [-3.49]	-0.082 [-3.54]
N	9572	9572	9572	9572	9572
R^2 (%)	0.261	1.022	0.686	0.863	0.714

Table 2: International Evidence

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

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

where $r_{m,t+1}^{country,e}$ is the futures return of one of the 10 largest equity markets: S&P 500 (United States, US), TOPIX (Japan, JP), CSI 300 (China, CN), FTSE 100 (Great Britain, GB), HSI (Hong Kong, HK), CAC 40 (France, FR), DAX (German, DE), TSX (Canada, CA), NIFTY 50 (India, IN), and SMI (Switzerland, CH). $s_t^{country}$ is the lagged difference between the equal-weight returns of the high-ownership and low-ownership stocks within the same country held by mutual funds. [Newey and West \(1987\)](#) t -statistics with 50 lags are reported. Coefficients that are significant at the 5% confidence level are in bold. The countries are ranked in descending order based on average total market value (MV). The sample periods are: 2001-2020 (US,JP,GB,HK,FR,DE,IN), 2010-2020 (CN), 2009-2020 (CA), 2004-2020 (CH).

country	$r_{m,t+1}^{US,e}$	$r_{m,t+1}^{JP,e}$	$r_{m,t+1}^{CN,e}$	$r_{m,t+1}^{GB,e}$	$r_{m,t+1}^{HK,e}$
	(1)	(2)	(3)	(4)	(5)
s_t	0.243 [3.14]	0.162 [2.53]	0.460 [3.79]	0.191 [2.47]	0.034 [0.53]
$r_{m,t}^e$	-0.115 [-3.39]	-0.134 [-4.71]	0.029 [0.85]	-0.027 [-1.41]	-0.062 [-2.11]
N	4999	4108	2266	4683	4245
R^2 (%)	1.501	1.304	0.747	0.121	0.285

country	$r_{m,t+1}^{FR,e}$	$r_{m,t+1}^{DE,e}$	$r_{m,t+1}^{CA,e}$	$r_{m,t+1}^{IN,e}$	$r_{m,t+1}^{CH,e}$
	(1)	(2)	(3)	(4)	(5)
s_t	0.130 [2.05]	0.039 [0.69]	0.190 [3.14]	0.071 [1.39]	0.536 [2.69]
$r_{m,t}^e$	-0.041 [-1.75]	-0.002 [-0.10]	-0.035 [-0.62]	-0.030 [-1.18]	-0.387 [-11.07]
N	4432	4642	2625	4536	3943
R^2 (%)	0.074	0.024	0.341	0.102	13.986

Table 3: Performance of a Daily Market-Timing Strategy

This table evaluates the performance of a daily out-of-sample market-timing strategy trading the S&P 500 futures. 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 five using our active ownership signal. The weight on the S&P 500 futures is bounded between -0.5 and 1.5. The out-of-sample equity premium is estimated with in-sample data since January 1983 until the date of portfolio formation; the out-of-sample volatility is estimated with a rolling window of 252 trading days. The sample period is from January 1983 to December 2020, and portfolio formation starts from January 1990. Panel A documents the key statistics of the trading strategy, where CER is 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 AO 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 the Fama-French five plus momentum factors. [White \(1980\)](#) heteroscedasticity-robust t -statistics are reported. Coefficients that are significant at the 5% confidence level are in bold.

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

$E(r_t^e)$ (%)	Std dev (%)	SR	Skewness	Kurtosis	CER (%)
14.70	16.16	0.91	0.76	38.32	6.71

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

α (%)	β_{MKT}	β_{SMB}	β_{HML}	β_{RMW}	β_{CMA}	β_{UMD}	IR
11.91 [4.35]	0.31 [8.32]						0.79
11.47 [4.13]	0.32 [8.35]	-0.12 [-3.36]	0.03 [0.66]			0.07 [2.68]	0.76
11.26 [4.05]	0.33 [7.69]	-0.11 [-3.18]	0.05 [0.97]	0.04 [0.88]	0.00 [0.02]	0.07 [2.65]	0.75

Table 4: Horizon of the S&P 500 Futures 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 of the table studies the S&P 500 futures return predictability within the five days since the generation of the signal. Panel B of the table studies the predictability of the intraday and overnight components of the daily S&P 500 futures return. Following [Bogousslavsky \(2021\)](#), we take the price at 9:45am as the open price to mitigate potential microstructure issues; for individual stock prices which we use to generate the signal, open price is defined as the midquote at 9:45am. $\{s_t\}$ 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 with 50 lags are reported. Coefficients that are significant at the 5% confidence level are in bold. The sample period for Panel A is from January 1983 to December 2020; the sample period for Panel B is from January 1998 to December 2020.

Panel A. Five-Day Market Predictability					
	$r_{m,t+1}^e$	$r_{m,t+2}^e$	$r_{m,t+3}^e$	$r_{m,t+4}^e$	$r_{m,t+5}^e$
	(1)	(2)	(3)	(4)	(5)
s_t	0.224 [4.72]	0.027 [0.50]	0.033 [1.00]	0.001 [0.04]	0.010 [0.27]
$r_{m,t}^e$	-0.094 [-3.83]	-0.063 [-2.81]	-0.063 [-2.84]	-0.062 [-2.86]	-0.063 [-2.86]
N	9571	9570	9569	9568	9567
R^2 (%)	1.021	0.379	0.385	0.370	0.371

Panel B. Intraday v.s. Overnight Market Predictability								
	$r_{m,t+1}^{e,ctc}$			$r_{m,t+1}^{e,cto}$			$r_{m,t+1}^{e,otc}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
s_t^{ctc}	0.213 [3.45]			0.077 [1.78]			0.139 [2.94]	
s_t^{cto}		0.112 [1.72]			0.040 [1.04]			0.044 [0.88]
s_t^{otc}			0.183 [2.71]			0.080 [1.55]		0.110 [2.03]
$r_{m,t}^e$	-0.106 [-2.95]	-0.088 [-2.67]	-0.090 [-2.71]	-0.027 [-1.62]	-0.021 [-1.51]	-0.022 [-1.49]	-0.081 [-3.45]	-0.067 [-2.87]
N	5668	5668	5668	5668	5668	5668	5668	5668
R^2 (%)	1.253	0.746	1.029	0.304	0.107	0.279	1.104	0.724

Table 5: Relation between Firm Characteristics and Active Ownership

This table reports the time-averages of firm characteristics across stocks sorted by active-mutual-fund ownership. The universe is limited to the stocks with market cap above the 20th percentile of NYSE stocks. The sample period is from January 1983 to December 2020 for market cap, bid-ask spread and Amihud illiquidity. The sample starts from January 1987 for analyst coverage and January 2001 for media coverage. Analyst coverage is measured as the number of analysts publish earnings forecasts for the company; media coverage is the number of news articles about the company published during the most recent three months prior to portfolio formation.

Active Ownership Quintile	1	2	3	4	5
Active Fund Ownership (%)	2.85	7.38	11.20	15.48	23.58
Market Capitalization (\$10 ⁹)	5.79	8.53	6.74	4.13	2.76
Bid-Ask Spread (%)	1.33	1.01	0.93	0.89	0.83
Amihud Illiquidity	0.21	0.05	0.04	0.03	0.03
Analyst Coverage	8.34	11.22	11.73	11.42	11.32
Media Coverage	52	71	62	48	38

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

This table presents the predictability of the daily S&P 500 futures return by the lagged signals extracted from alternative firm characteristics:

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

where $r_{m,t+1}^e$ is the S&P 500 futures return on date $t+1$; s_t is the difference between the equal-weighted average of the high-active-ownership stock returns and the low-active-ownership stock returns on date t ; $s_t^{characteristic}$ 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 cap above the 20th percentile of NYSE stocks. Analyst coverage is from the IBES. Media coverage is from Ravenpack. [Newey and West \(1987\)](#) t -statistics with 50 lags are reported. Coefficients that are significant at the 5% confidence level are in bold. The sample period for active mutual fund ownership, market cap, bid-ask spread, Amihud illiquidity is from January 1983 to December 2020; the sample period for analyst coverage is from January 1987 to December 2020; and the sample period for media coverage is from January 2001 to December 2020.

$s_t^{characteristic}$	$r_{m,t+1}^e$					
	(1)	(2)	(3)	(4)	(5)	(6)
Active Fund Ownership	0.220 [4.53]	0.216 [4.40]	0.210 [4.22]	0.202 [3.78]	0.232 [3.10]	0.218 [2.64]
Market Capitalization	0.021 [0.34]					-0.209 [-1.94]
Bid-Ask Spread		-0.021 [-0.78]				-0.017 [-0.29]
Amihud Illiquidity			-0.037 [-0.66]			-0.113 [-0.70]
Analyst Coverage				0.055 [0.68]		-0.105 [-1.23]
Media Coverage					0.018 [0.22]	0.174 [1.39]
$r_{m,t-1}^e$	-0.095 [-3.94]	-0.095 [-3.87]	-0.098 [-4.07]	-0.097 [-3.84]	-0.116 [-3.15]	-0.136 [-2.92]
N	9572	9572	9572	8563	5025	5025
R^2 (%)	1.025	1.028	1.051	1.081	1.453	1.800

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

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

$$r_{m,t+1}^e = a_0 + a_1 s_t^{institution} + a_3 r_{m,t}^e + \epsilon_{t+1},$$

where $r_{m,t+1}^e$ is the S&P 500 futures return; $s_t^{institution}$ is the lagged difference between the equal-weighted average returns of the high-ownership and low-ownership stocks held by a specific type of financial institution. The signals are 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. Mutual fund holdings and institution holdings are from the Thomson Reuters mutual fund holdings and the 13F holdings data, respectively. Hedge fund holdings are from the FactSet Global Ownership data. Institution classification follows [Kojen and Yogo \(2019\)](#). [Newey and West \(1987\)](#) t -statistics 50 lags are reported. Coefficients that are significant at the 5% confidence level are in bold. The sample period is from January 1988 to December 2020 for passive fund and ETF; from April 1999 to June 2020 for hedge fund; and from January 1983 to December 2020 for the rest of the institutions.

$s_t^{institution}$	$r_{m,t+1}^e$								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Active Mutual Fund	0.224 [4.72]							0.249 [4.18]	0.252 [3.18]
Passive Fund and ETF		0.062 [1.63]						0.084 [1.12]	0.134 [1.51]
Investment Advisor			0.147 [2.73]					0.006 [0.11]	-0.162 [-1.95]
Pension Fund				0.124 [1.74]				0.117 [1.32]	0.149 [1.36]
Bank					0.045 [0.86]			-0.045 [-0.70]	-0.016 [-0.23]
Insurance Company						0.073 [1.33]		-0.246 [-2.54]	-0.268 [-2.12]
Hedge Fund							0.158 [2.36]		0.218 [2.27]
$r_{m,t}^e$	-0.094 [-3.83]	-0.080 [-3.42]	-0.075 [-3.17]	-0.068 [-3.32]	-0.062 [-2.74]	-0.067 [-3.21]	-0.100 [-2.97]	-0.110 [-4.23]	-0.128 [-3.80]
N	9572	8310	9572	9572	9572	9572	5347	8310	5347
R^2 (%)	1.022	0.660	0.608	0.619	0.417	0.439	1.029	1.447	1.890

Table 8: Predictability of Aggregate Stock News Sentiment

Panel A of this table presents the predictability of the aggregate stock news sentiment by the lagged active ownership signal:

$$css_{t+1} = a_0 + a_1 s_t + a_3 r_{m,t}^e + \epsilon_{t+1},$$

where css_{t+1} is the average Ravenpack Composite Sentiment Score of the business-related news across all firms on date $t + 1$ weighted by market capitalization; s_t is the difference between the equal-weighted average returns of the high-active-ownership and low-active-ownership stocks on date t ; $r_{m,t}^e$ is the S&P 500 futures return on date t . The first column includes all stocks being traded in the market, whereas the second column limits the sample to the S&P 500 firms. The construction of the aggregate stock news sentiment includes all news articles except those belong to the "stock-prices" topic group. Panel B of the table presents the contemporaneous regression of the S&P 500 futures return on the predicted aggregate stock news sentiment. $\{s_t\}$ 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 with 50 lags are reported. Coefficients that are significant at the 5% confidence level are in bold. The sample period is from January 2001 to December 2020.

Panel A. Predicting Future Aggregate News Sentiment with Signal

	css_{t+1}	$css_{t+1}^{S\&P500}$
	(1)	(2)
s_t	3.658 [2.10]	3.702 [2.00]
$r_{m,t}^e$	1.767 [2.72]	1.896 [2.67]
N	5026	5026
R^2 (%)	0.260	0.275

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

	$r_{m,t+1}^e$	
	(1)	(2)
$\widehat{css}_{t+1}$	0.066 [3.13]	
$\widehat{css}_{t+1}^{S\&P500}$		0.065 [3.13]
$r_{m,t}^e$	-0.232 [-3.51]	-0.239 [-3.51]
N	5026	5026
R^2 (%)	1.492	1.492

Table 9: Daily S&P 500 Futures Predictability by Subgroups of Active Mutual Funds

This table presents the predictability of the daily S&P 500 futures return by the lagged signals extracted from the stocks owned by high or low information ratio (or turnover) mutual funds:

$$r_{m,t+1}^e = a_0 + a_1 s_t^{highIR} + a_2 s_t^{lowIR} + a_3 r_{m,t}^e + \epsilon_{t+1},$$

$$r_{m,t+1}^e = a_0 + a_1 s_t^{highTO} + a_2 s_t^{lowTO} + a_3 r_{m,t}^e + \epsilon_{t+1},$$

where $r_{m,t+1}^e$ is the excess S&P 500 futures return; s_t^{highIR} (s_t^{lowIR}) is the lagged difference between the equal-weighted average returns of the high-ownership and low-ownership stocks held by high (low) information ratio funds; s_t^{highTO} (s_t^{lowTO}) is the lagged difference between the equal-weighted average returns of the high-ownership and low-ownership stocks held by high (low) turnover funds. The information ratios of the fund are measured using [Carhart \(1997\)](#) 4-factor model during the 24 months prior to signal construction. The turnover ratio is from the CRSP Mutual Fund Databased directly, 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. The signals are 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 with 50 lags are reported. Coefficients that are significant at the 5% confidence level are in bold. The sample period is from January 1983 to December 2020.

s_t^{type}	$r_{m,t+1}^e$					
	(1)	(2)	(3)	(4)	(5)	(6)
high IR	0.231 [4.55]		0.218 [4.38]			
Low IR		0.095 [2.39]	0.045 [1.16]			
High Turnover				0.195 [4.79]		0.188 [4.76]
Low Turnover					0.121 [2.49]	0.074 [1.94]
$r_{m,t}^e$	-0.090 [-4.16]	-0.075 [-3.27]	-0.094 [-4.15]	-0.107 [-4.30]	-0.065 [-2.97]	-0.108 [-4.28]
N	9318	9318	9318	9318	9318	9318
R^2 (%)	1.066	0.522	1.086	1.068	0.578	1.207

Table 10: Daily Industry Return Predictability

This table presents the predictability of the daily industry returns (in excess of the market return) on the lagged signals extracted from the stocks owned 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 market return; $s_t^{industry}$ 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 the all-but-micro-cap stocks, defined as the stocks with market cap above the 20th percentile of NYSE stocks. The market return is the return of the value-weighted market index. Industry classification follows [Kacperczyk et al. \(2005\)](#). [Newey and West \(1987\)](#) t -statistics with 50 lags are reported. Coefficients that are significant at the 5% confidence level are in bold. The industries are ranked in descending order based on the average ownership by active mutual funds of the stocks within. The sample period is from January 1983 to December 2020.

	$\tilde{r}_{t+1}^{manftr}$	$\tilde{r}_{t+1}^{bussv}$	$\tilde{r}_{t+1}^{fin}$	$\tilde{r}_{t+1}^{hlthcr}$	$\tilde{r}_{t+1}^{whlsl}$	$\tilde{r}_{t+1}^{telcm}$	$\tilde{r}_{t+1}^{engy}$	$\tilde{r}_{t+1}^{nondur}$	$\tilde{r}_{t+1}^{dur}$	$\tilde{r}_{t+1}^{util}$
$s_t^{industry}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Manufacturing	0.039 [2.74]									
Business Service		0.062 [3.80]								
Finance			0.041 [2.37]							
Healthcare				0.015 [1.73]						
Wholesale					-0.003 [-0.24]					
Telecom						-0.004 [-0.69]				
Energy							0.017 [0.85]			
Non-durable								0.005 [0.53]		
Durable									-0.009 [-1.17]	
Utility										-0.006 [-0.22]
$\tilde{r}_t^{industry}$	0.121 [7.04]	0.051 [3.05]	-0.021 [-0.79]	0.073 [4.71]	0.042 [2.52]	0.050 [3.10]	0.042 [2.68]	0.043 [2.38]	-0.004 [-0.23]	0.077 [4.45]
N	9572	9572	9572	9572	9572	9572	9572	9572	9572	9572
R^2 (%)	1.824	0.923	0.162	0.606	0.151	0.231	0.236	0.160	0.007	0.561
AO (%)	12.11	11.72	10.52	10.26	9.82	9.18	8.78	8.77	8.42	8.09

Table 11: Cross-Predictability of Daily Stock Returns by Active Mutual Fund Ownership

This table presents the lead-lag relation of the daily returns among the 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. Following [Bogousslavsky \(2021\)](#), the open price of an individual stock is defined as the midquote at 9:45am to mitigate potential microstructure issues. The universe of the all-but-micro-cap stocks, defined as the stocks with market cap above the 20th percentile of NYSE stocks, are divided into five groups with equal number of stocks by active mutual fund ownership. Equal-weighted and value-weighted returns are computed for each group of stocks. [Newey and West \(1987\)](#) t -statistics with 50 lags are reported. Coefficients that are significant at the 5% confidence level are in bold. The intraday stock transactions data is from NYSE Trade and Quote (TAQ) database. The sample period is from January 1993 to December 2020.

<i>Panel A. Equal-Weighted Portfolios Formed by Active Mutual Fund Ownership</i>					
	$r_{t+1}^{h,ew}$	$r_{t+1}^{4,ew}$	$r_{t+1}^{3,ew}$	$r_{t+1}^{2,ew}$	$r_{t+1}^{l,ew}$
	(1)	(2)	(3)	(4)	(5)
$r_t^{h,ew}$	0.398	0.341	0.320	0.323	0.333
	[4.85]	[4.27]	[3.97]	[4.04]	[4.13]
$r_t^{4,ew}$	-0.034	-0.011	-0.003	-0.019	-0.058
	[-0.29]	[-0.10]	[-0.03]	[-0.17]	[-0.48]
$r_t^{3,ew}$	-0.051	-0.043	-0.097	-0.181	-0.222
	[-0.41]	[-0.36]	[-0.79]	[-1.36]	[-1.52]
$r_t^{2,ew}$	-0.305	-0.313	-0.248	-0.190	-0.162
	[-2.21]	[-2.30]	[-1.79]	[-1.25]	[-1.05]
$r_t^{l,ew}$	-0.067	-0.026	-0.015	0.023	0.083
	[-0.84]	[-0.33]	[-0.20]	[0.28]	[1.09]
N	7023	7022	7022	7022	7022
R^2 (%)	1.468	1.234	1.023	0.916	0.729

Panel B. Value-Weighted 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)
$r_t^{h,vw}$	0.126 [2.76]	0.113 [2.69]	0.097 [2.15]	0.101 [2.46]	0.099 [2.85]
$r_t^{4,vw}$	-0.045 [-0.63]	-0.043 [-0.69]	-0.019 [-0.34]	-0.038 [-0.76]	-0.058 [-1.16]
$r_t^{3,vw}$	0.028 [0.48]	0.007 [0.12]	0.001 [0.01]	-0.008 [-0.16]	-0.032 [-0.69]
$r_t^{2,vw}$	-0.096 [-1.88]	-0.060 [-1.27]	-0.054 [-1.28]	-0.095 [-2.13]	-0.081 [-1.90]
$r_t^{l,vw}$	-0.068 [-1.89]	-0.079 [-2.20]	-0.089 [-2.57]	-0.031 [-1.12]	0.020 [0.55]
N	7023	7022	7022	7022	7022
R^2 (%)	0.483	0.557	0.518	0.629	0.690

Appendix A. Proof of Proposition 1

Under the equilibrium, the demand for risky asset of the informed investor is

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

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,

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

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

and

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

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

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

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

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

The risky demand of the uninformed investor is

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

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

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

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

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

where

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

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

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:

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

where the parameters solve

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

Verification of the value function:

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

where

$$\begin{aligned}
\vec{X}_t &= \frac{1}{a} \left[\Sigma \left(\vec{P}_{t+1} + \vec{D}_{t+1} | \mathcal{F}_t \right) \right]^{-1} \left[\mathbb{E} \left(\vec{P}_{t+1} + \vec{D}_{t+1} | \mathcal{F}_t \right) - R \vec{P}_t \right] \\
c &= \vec{X}_t' \mathbb{E} \left(\vec{P}_{t+1} + \vec{D}_{t+1} | \mathcal{F}_t \right) - \frac{a}{2} \vec{X}_t' \Sigma \left(\vec{P}_{t+1} + \vec{D}_{t+1} | \mathcal{F}_t \right) \vec{X}_t + \vec{X}_t' \cdot \vec{P}_t R \\
C_t^* &= \frac{-\log \frac{a\beta R}{\alpha} + aRW_t - ac + b}{\alpha + aR} \\
&= \frac{-\log \frac{a\beta R}{\alpha} - ac + b}{\alpha + aR} + \frac{aR}{\alpha + aR} W_t \\
&\equiv p + qW_t
\end{aligned}$$

and a, b satisfy:

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

Appendix B. Robustness Checks

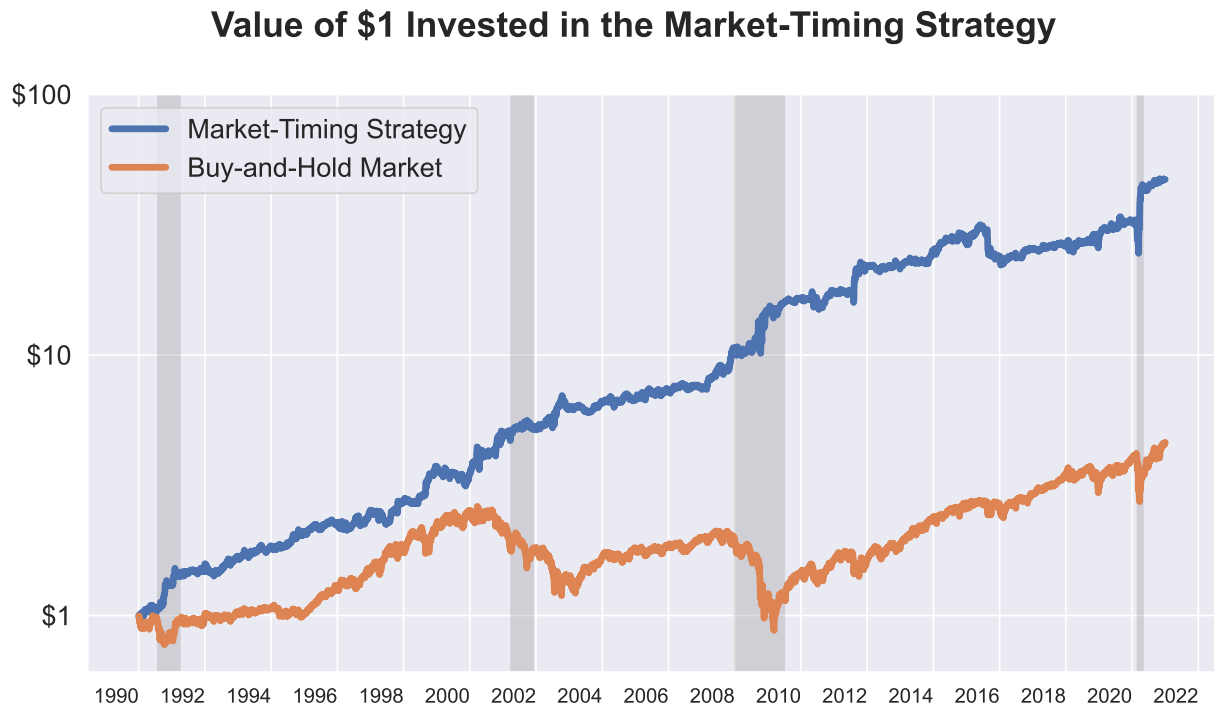


Fig. B1. Performance of a Daily Market-Timing Strategy: Signals Extracted From the 2-Month Lagged Active Ownership

The figure plots the (log) cumulative performance of our daily market-timing strategy trading the S&P 500 futures. 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 five using our active-ownership signal. The weight on the S&P 500 futures is bounded between -0.5 and 1.5. The out-of-sample equity premium is estimated with in-sample data since January 1983 until the date of portfolio formation; the out-of-sample volatility is estimated with a rolling window of 252 trading days. 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 January 1983 to December 2020, 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:

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

where $r_{m,t+1}^e$ 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 Index (S&P500), Fama-French Market (FF); 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 with 50 lags are reported. Coefficients that are significant at the 5% confidence level are in bold. The sample period is from February 1993 to December 2020.

	$r_{FUT,t+1}^e$	$r_{EMini,t+1}^e$	$r_{SPDR,t+1}^e$	$r_{S\&P500,t+1}^e$	$r_{VW,t+1}^e$
	(1)	(2)	(3)	(4)	(5)
s_t	0.199 [3.72]	0.202 [3.17]	0.196 [3.58]	0.211 [3.71]	0.229 [3.96]
$r_{m,t}^e$	-0.103 [-3.68]	-0.113 [-3.04]	-0.111 [-3.61]	-0.121 [-3.38]	-0.100 [-2.73]
R^2 (%)	7025	5611	7025	7025	7025
N	1.211	1.405	1.337	1.574	1.301

Table B2: Signals Extracted From the 2-Month Lagged Active Ownership

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.

	$r_{m,t+1}^e$				
	(1)	(2)	(3)	(4)	(5)
s_t	0.126 [3.69]	0.208 [4.30]			
$Sign(s_t)$			0.055 [3.71]		
s_t^+				0.296 [4.02]	
s_t^-					0.235 [3.40]
$r_{m,t}^e$		-0.090 [-3.63]	-0.074 [-3.21]	-0.082 [-3.41]	-0.078 [-3.34]
N	9578	9578	9578	9578	9578
R^2 (%)	0.214	0.913	0.563	0.825	0.625
R_{OOS}^2 (%)	0.210***	0.927***	0.605***	0.751***	0.528***

Table B3: Performance of a Daily Market-Timing Strategy: Signals Extracted From the 2-Month Lagged Active Ownership

This table evaluates the performance of a daily out-of-sample market-timing strategy trading the S&P 500 futures. 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 five using our active ownership signal. The weight on the S&P 500 futures is bounded between -0.5 and 1.5. The out-of-sample equity premium is estimated with in-sample data since January 1983 until the date of portfolio formation; the out-of-sample volatility is estimated with a rolling window of 252 trading days. The sample period is from January 1983 to December 2020, and the portfolio formation starts at January 2, 1990. Panel A documents the key statistics of the trading strategy, where CER is 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 AO 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 the Fama-French five plus momentum factors. [White \(1980\)](#) heteroscedasticity-robust t -statistics are reported. Coefficients that are significant at the 5% confidence level are in bold.

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

$E(r_t^e)$ (%)	Std dev (%)	SR	Skewness	Kurtosis	CER (%)
13.70	15.97	0.86	1.15	48.87	5.88

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

α (%)	β_{MKT}	β_{SMB}	β_{HML}	β_{RMW}	β_{CMA}	β_{UMD}	IR
10.93 [4.04]	0.31 [8.28]						0.73
10.43 [3.82]	0.32 [8.35]	-0.11 [-3.03]	0.04 [1.09]			0.07 [2.89]	0.70
10.04 [3.67]	0.33 [7.84]	-0.10 [-2.74]	0.05 [1.05]	0.07 [1.42]	0.03 [0.51]	0.07 [2.75]	0.68

Table B4: International Evidence: Signals Extracted From the 2-Month Lagged Active Ownership

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

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

where $r_{m,t+1}^{country,e}$ is the futures return of one of the 10 largest equity markets: S&P 500 (United States, US), TOPIX (Japan, JP), CSI 300 (China, CN), FTSE 100 (Great Britain, GB), HSI (Hong Kong, HK), CAC 40 (France, FR), DAX (German, DE), TSX (Canada, CA), NIFTY 50 (India, IN), and SMI (Switzerland, CH). $s_t^{country}$ is the lagged difference between the equal-weight returns of the high-ownership and low-ownership stocks within the same country held by mutual funds. [Newey and West \(1987\)](#) t -statistics with 50 lags are reported. Coefficients that are significant at the 5% confidence level are in bold. The countries are ranked in descending order based on average total market value (MV). The sample periods for the 10 markets, which depend on the availability of the data of market futures, stock returns, and institutional ownership, are 2001-2020 (US,JP,GB,HK,FR,DE,IN), 2010-2020 (CN), 2009-2020 (CA), 2004-2020 (CH).

country	$r_{m,t+1}^{US,e}$	$r_{m,t+1}^{JP,e}$	$r_{m,t+1}^{CN,e}$	$r_{m,t+1}^{GB,e}$	$r_{m,t+1}^{HK,e}$
	(1)	(2)	(3)	(4)	(5)
s_t	0.215 [2.69]	0.184 [2.91]	0.586 [3.55]	0.244 [3.04]	0.037 [0.59]
$r_{m,t}^e$	-0.109 [-3.28]	-0.138 [-4.73]	0.027 [0.92]	-0.032 [-1.68]	-0.068 [-2.37]
N	4999	4109	2309	4690	4253
R^2 (%)	1.351	1.361	1.141	0.226	0.346

country	$r_{m,t+1}^{FR,e}$	$r_{m,t+1}^{DE,e}$	$r_{m,t+1}^{CA,e}$	$r_{m,t+1}^{IN,e}$	$r_{m,t+1}^{CH,e}$
	(1)	(2)	(3)	(4)	(5)
s_t	0.150 [2.21]	0.008 [0.16]	0.165 [2.76]	0.025 [0.54]	0.589 [2.84]
$r_{m,t}^e$	-0.045 [-1.95]	0.005 [0.22]	-0.032 [-0.57]	-0.032 [-1.30]	-0.390 [-11.07]
N	4439	4649	2662	4543	3986
R^2 (%)	0.111	0.037	0.242	0.072	14.237

Table B5: Robustness: Alternative Financial Institutions, 2001 - 2020

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

$$r_{m,t+1}^e = a_0 + a_1 s_t^{institution} + a_3 r_{m,t}^e + \epsilon_{t+1},$$

where $r_{m,t+1}^e$ is the S&P 500 futures return; $s_t^{institution}$ is the lagged difference between the equal-weighted average returns of the high-ownership and low-ownership stocks held by a specific type of financial institution. The signals are 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. Mutual fund holdings and institution holdings are from the Thomson Reuters mutual fund holdings and the 13F holdings data, respectively. Hedge fund holdings are from the FactSet Global Ownership data. Institution classification follows [Kojen and Yogo \(2019\)](#). [Newey and West \(1987\)](#) t -statistics 50 lags are reported. Coefficients that are significant at the 5% confidence level are in bold. The sample period is from January 2001 to December 2020.

	$r_{m,t+1}^e$								
$s_t^{institution}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Active Mutual Fund	0.243 [3.14]							0.277 [3.37]	0.249 [2.99]
Passive Fund and ETF		0.048 [0.78]						0.143 [1.65]	0.157 [1.73]
Investment Advisor			0.136 [1.61]					0.046 [0.59]	-0.116 [-1.23]
Pension Fund				0.017 [0.20]				0.127 [1.15]	0.140 [1.18]
Bank					-0.069 [-1.21]			-0.060 [-0.77]	0.017 [0.21]
Insurance Company						-0.052 [-0.72]		-0.322 [-2.56]	-0.314 [-2.37]
Hedge Fund							0.174 [2.57]		0.208 [1.99]
$r_{m,t}^e$	-0.115 [-3.41]	-0.097 [-3.19]	-0.096 [-3.08]	-0.095 [-3.12]	-0.097 [-3.18]	-0.095 [-3.16]	-0.113 [-3.31]	-0.124 [-3.95]	-0.139 [-3.98]
N	4905	4905	4905	4905	4905	4905	4905	4905	4905
R^2 (%)	1.501	0.898	1.049	0.865	0.926	0.891	1.344	1.924	2.114