Do Mutual Fund Managers Care About Star Ratings? Evidence from Portfolio Pumping*

Sanghyun (Hugh) Kim

October 1, 2019

* Hugh Kim (hugh.kim@utdallas.edu) is at the University of Texas at Dallas, 800 W. Campbell Road, Richardson, TX 75080. I wish to thank my adviser, Vikram Nanda, for his helpful comments, patience, and invaluable advice throughout all stages of my doctoral dissertation research, and Kelsey Wei, Umit Gurun, Alessio Saretto, Qinghai Wang, Munhee Han, Harold Zhang, Nina Baranchuk, Steven Xiao, Yexiao Xu, Feng Zhao, and seminar participants at the University of Texas at Dallas for helpful comments.

Do Mutual Fund Managers Care About Star Ratings? Evidence from Portfolio Pumping

Abstract

This paper reveals that the discrete nature of Morningstar ratings gives mutual fund managers powerful incentives to inflate their month-end performance. Compared to their distant peers, mutual funds near a rating threshold experience substantially larger gains on the last trading day of the month, which partially dissipates on the next trading day. This effect is more pronounced among funds with a greater incentive and ability to pump up their portfolios. In addition, stocks predominantly held by funds close to a rating cutoff also earn significantly higher returns at the month-end, especially during the last minutes of the trading session. Less liquid stocks are naturally more exposed to pumping. Placebo tests exploiting a change in the Morningstar rating methodology around June 2002 provide corroborating evidence that the threshold effect of star ratings on portfolio pumping is likely causal.

Keywords: Morningstar ratings, mutual funds, portfolio pumping, price manipulation, threshold effects

1 Introduction

Since its introduction in 1985, Morningstar's five-star rating system has become widely accepted in the mutual fund industry.[1] Unlike well-known performance measures commonly used in academia (e.g., Jensen (1969), Carhart (1997), Daniel et al. (1997)), star ratings offer less sophisticated investors a simple and intuitive tool to use to allocate their capital across mutual funds. Del Guercio and Tkac (2008) show that discrete star ratings have a powerful influence on investor flows, independent of continuous performance rankings. Recently, Ben-David et al. (2019) have forcibly argued that star ratings are the main determinant of capital allocation across mutual funds.[2] Given that mutual fund investors rely heavily on star ratings, fund managers might also be pressured to maintain desirable star ratings.

In this paper, I present evidence that discrete Morningstar ratings give mutual fund managers powerful incentives to inflate their month-end performance when their funds are likely to finish the month in the vicinity of a rating threshold. Compared to their distant peers, mutual funds near a rating threshold experience substantially larger gains on the last trading day of the month, about a third of which dissipates on the next trading day. Stocks widely held by funds close to a rating threshold also earn significantly higher returns at the month-end, especially during the last minutes of the trading session. This end-of-the-month return pattern is consistent with a practice known as "portfolio pumping" (Zweig (1997), Carhart et al. (2002)). Since open-end mutual funds calculate their net asset values (NAVs) from the closing prices of their holdings, mutual fund managers can "pump up" the prices of their holdings through aggressive trading of stocks they already own.

This paper sheds new light on incentives that mutual fund managers face and how these incentives can distort asset prices. The literature often relies on the convex flow–performance relation (Ippolito

[1] Mutual funds are rated every month on a scale of one star (lowest) to five stars (highest) on the basis of Morningstar's proprietary algorithm known as "Risk-Adjusted Returns" (RAR). On the basis of percentile rankings of Morningstar RAR, the top 10% of funds receive five stars, the next 22.5% four stars, the middle 35% three stars, the next 22.5% two stars, and the bottom 10% one star.

[2] The idea that mutual fund investors use risk-adjusted returns in their capital allocation decision is appealing and has motivated some prominent financial economists to use mutual fund flows to test asset pricing models. Berk and van Binsbergen (2016), for instance, conclude that the Capital Asset Pricing Model (CAPM) is the "closest to the asset pricing model investors are actually using". Employing a different methodology, Barber, Huang, and Odean (2016) also reach the same conclusion. In a recent study, however, Ben-David et al. (2019) show that simple and readily available star ratings explain mutual fund investors' behavior much better than any asset pricing model.

(1992), Sirri and Tufano (1998)) to explain how mutual fund managers respond to incentives in various settings. Carhart et al. (2002) propose the "leaning for the tape" hypothesis (picture a runner at the finish line) as an explanation for widespread NAV inflation of mutual funds at the quarter-end (especially the year-end). However, since portfolio pumping is an illegal practice, even if some fund managers pump up their portfolios, they would likely take caution. Hu et al. (2014), for instance, find that year-end price inflation derives from depressed institutional selling rather than aggressive institutional buying. In this paper, I show that pumping managers tend to inflate their month-end performance when even a small difference in performance rankings can lead to a discrete change in star ratings and to a large difference in future investor flows.[3]

I further exploit a major change in the Morningstar rating methodology in June 2002 to provide corroborating evidence that the threshold effect of star ratings on portfolio pumping is likely causal. Morningstar classifies mutual funds into finer categories (e.g., U.S. Equity Large Growth) within a broader category group (e.g., U.S. Equity). For open-end U.S. equity funds, Morningstar assigns each fund to one of its nine (3×3) categories based on its size tilt (small, mid-cap, or large) and value tilt (value, blend, or growth). While all U.S. equity funds were ranked against each other until May 2002, Morningstar started raking U.S. equity funds within one of its 3×3 categories in June 2002. I test for placebo effects by reversing this June 2002 change in the rating methodology[4] and find that *placebo* percentile rankings around rating thresholds have null effects on portfolio pumping.

My primary source of data is Morningstar Direct, a survivorship-bias-free research platform offered by Morningstar Inc., for the period from 1990 to 2018. To examine whether star ratings incentivize fund managers to inflate their month-end performance, I measure the distance between a fund's percentile ranking and its nearest rating threshold on the second-to-last trading day of each month. Star ratings are determined by Morningstar's "Risk-Adjusted Returns" (RAR) over the prior 3, 5, and 10 years, depending

[3] Reuter and Zitzewitz (2015) estimate that about 22% of the difference in future fund flows received by five-star and one-star funds represents a causal effect of the difference in star ratings on investor flows.

[4] That is, for placebo percentile rankings, I use within-category percentile rankings until May 2002 and group percentile rankings starting in June 2002. Although within-category percentile rankings and group percentile rankings are highly correlated, the correlation between their *distances* to the nearest rating thresholds (10^{th}, 32.5^{th}, 77.5^{th}, or 90^{th} percentiles) is quite low.

on data availability. In this paper, I focus on the three-year rating, which is the shortest track record and the easiest one to influence. To obtain percentile rankings that are not contaminated by the outcomes of the last trading session of the month, while replicating Morningstar RAR percentile rankings as closely as possible, I rank funds on the basis of the Sharpe ratio calculated using daily returns in excess of the risk-free rate over the prior three years, but only through the second-to-last trading day of the month.[5] Following Morningstar, I rank all U.S. equity funds against each other until May 2002, while ranking funds within each of Morningstar's 3×3 categories starting in June 2002.

First, I present fund-level evidence that some fund managers inflate their month-end NAVs when their funds are likely to finish the month in the vicinity of a rating threshold. Specifically, the (squared) distance to a rating threshold on the second-to-last trading day of the month negatively predicts the fund return on the last trading day of the month while positively predicting the fund return on the first trading day of the subsequent month. Compared to funds that are about 10^{th} percentile away, funds near a rating threshold on average earn 144–167 basis points higher annualized returns on the last trading day of the month.[6] In placebo tests in which I rank funds by reversing the June 2002 rating methodology change, the relation between the distance to a rating threshold and the fund return around the turn of the month disappears.

In the cross-section of mutual funds, the threshold effect of star ratings on month-end NAV inflation is more pronounced among funds that are better able to pump up their portfolios. Carhart et al. (2002) argue that if fund managers are indeed "marking up" their funds' quarter-end (or year-end) NAVs, funds with the greatest ability to influence their performance are most likely to pump. For instance, the closing prices of less liquid stocks would presumably be easier to influence. Consistent with portfolio pumping, I find that the threshold effect of star ratings on month-end NAV inflation is much greater among small- and

[5] Sharpe (1998) reports that, in his sample of 1,286 equity funds from 1994 through 1996, the correlation between Morningstar RAR percentile and the Sharpe ratio percentile was 0.986. He notes that this period is characterized by high returns that contribute to a good fit.

[6] To put the numbers in perspective, Carhart et al. (2002) find that among nine Lipper mutual fund indexes, the abnormal return on the last trading day of the year ranges from 174 basis points per year for the small-cap growth fund index to 25 basis points per year for the large-cap value fund index during the period from July 13, 1992 to July 7, 2000.

mid-cap funds. The cross-sectional results corroborate that the relation between the distance to a rating threshold and month-end NAV inflation is indeed driven by portfolio pumping activities.

Next, I provide more direct evidence regarding the threshold effect of star ratings on portfolio pumping using Thomson-Reuters' Mutual Fund Ownership data and NYSE's intraday Trade and Quote (TAQ) data. Although star ratings are updated every month and portfolio pumping can occur at any month-end, I focus on calendar quarter-ends in order to use the most accurate snapshot of stock holdings of mutual funds (see, e.g., Ben-David et al. (2013)). I find that stocks with higher ownership by funds near a rating threshold and lower ownership by funds distant from rating thresholds have significantly higher returns on the last trading day of the month, especially during the last 30 minutes before the close. Again, the relation between month-end price inflation and relative ownership by funds near a rating threshold disappears in placebo tests, mitigating concerns about endogenous mutual fund ownership. In the cross-section of stocks, the threshold effect of star ratings on price inflation is more pronounced among less liquid stocks.

This paper contributes to three strands of literature. The first strand documents strong evidence of portfolio pumping activities at quarter-ends (especially year-ends) and typically by top-performing funds (Carhart et al. (2002), Ben-David et al. (2013), Hu et al. (2014)). This paper contributes to this literature by presenting strong evidence of portfolio pumping at all month-ends and across a wide spectrum of performance rankings, with sharper identification using rating thresholds. The second strand of literature finds that Morningstar ratings have a powerful influence on investor flows, independent of continuous performance rankings (Del Guercio and Tkac (2008), Reuter and Zitzewitz (2015)). In a recent study, Ben-David et al. (2019) show that simple and readily available star ratings explain mutual fund investors' behavior much better than any asset pricing models. This paper complements this literature by taking the perspective of fund managers and showing that Morningstar ratings give mutual fund mangers powerful incentives to inflate their performance when their funds move closer to a rating threshold. The third strand of literature finds strong evidence of price manipulation around various thresholds. For instance, Ni, Pearson, and Poteshman (2005) report that stock prices tend to cluster around option strike prices on expiration dates.

The remainder of this paper is organized as follows. In the next section, I provide institutional details on Morningstar ratings. Section 3 introduces my data sets and describes how I construct variables used in the subsequent analysis. Sections 4 and 5 present evidence on the threshold effect of star ratings on portfolio pumping at the fund-level and at the holding-level, respectively. Section 6 concludes.

2 Morningstar Ratings

Since the introduction of its five-star rating system in 1985, Morningstar has become the undisputed leader of the fund rating industry (Del Guercio and Tkac (2008)). Morningstar ratings have been shown to have a strong independent influence on investor flows (Del Guercio and Tkac (2008), Reuter and Zitzewitz (2015), Ben-David et al. (2019)).

Mutual fund share classes are rated by Morningstar on a monthly basis and on an integer scale of one star (the lowest rating) to five stars (the highest rating). Table 1 shows the transition matrix where each element in row i and column j shows the probability of a mutual fund share class receiving rating j in month t conditional on its receiving rating i in month $t-1$. There is considerable time-series variation in star ratings and more than 10% of funds tend to experience changes in their star ratings each month.

[Insert Table 1]

The star ratings assigned to each mutual fund share class are determined by its (within-category) rankings of Morningstar "Risk-Adjusted Return" (RAR), which adjusts for risk and accounts for all sales charges, over the prior three, five, and ten years, depending on the data availability. Share classes less than three years old are not rated and the overall rating is determined by the weighted average of three-, five-, and ten-year ratings, depending on the age of the share class. This study focuses on the three-year rating, which is the shortest track record and the easiest one to influence.

Morningstar classifies mutual funds into finer categories (e.g., U.S. Equity Large Growth) within a broader category group (e.g., U.S. Equity). This study focuses on open-end U.S. equity funds, for which Morningstar assigns each fund to one of its nine (3×3) categories based on its size tilt (Small, Mid-Cap, or Large) and value tilt (Value, Blend, or Growth). Prior to June 2002, all U.S. equity funds were ranked against each other and since then Morningstar has started ranking funds within each category. On the basis of (within-category) percentile rankings of Morningstar RAR, the top 10% of funds receive five stars, the next 22.5% four stars, the middle 35% three stars, the next 22.5% two stars, and the bottom 10% receive one star. As a result, small differences in past performance, such as going from the 89th percentile to 90th percentile, lead to discrete changes in Morningstar ratings, such as going from four stars to five stars (see Reuter and Zitzewitz (2015) for detailed discussions on the discrete nature of Morningstar ratings).

3 Data and Variable Construction

3.1 Fund-Level Variables

My primary data come from Morningstar Direct, from which I obtain data on daily returns, Morningstar categories, star ratings, inception dates, and index-fund indicators. Morningstar ratings are updated every month and based on Morningstar "Risk-Adjusted Return" (RAR), which is calculated using monthly returns. In order to obtain percentile rankings that are not contaminated by the outcomes of the last trading session of the month, I use daily returns over the prior three years, but only through the second-to-last trading day of the month. In order to closely replicate percentile rankings based on Morningstar RAR, I calculate the Sharpe ratio of each mutual fund share class every month using daily returns in excess of the risk-free rate over the prior three years excluding the last trading day. Sharpe (1998) shows that ranking funds based on Morningstar RAR gives results that are similar to ranking funds on the basis of Sharpe ratio.[7]

There are four rating thresholds dividing all funds into five star ratings: 10th, 32.5th, 77.5th, and 90th percentiles. The distance to a rating threshold is calculated for each share class every month by the distance between its Sharpe ratio (within-category) percentile ranking and its nearest threshold on the second-to-last trading day of the month. Following the Morningstar rating methodology, I rank all U.S.

[7] Sharpe (1998) reports that, in his sample of 1,286 equity funds from 1994 through 1996, the correlation between Morningstar RAR percentile and the Sharpe ratio percentile was 0.986.

equity funds against each other until May 2002 and since then funds are ranked within each of Morningstar's 3×3 categories starting in June 2002. Although my study focuses on actively-managed funds, I keep index funds in the ranking procedure in line with the Morningstar rating methodology.

In some of my empirical analyses, I conduct placebo tests in order to provide corroborating evidence that the threshold effects that I document in this paper are likely causal. To accomplish this, I calculate placebo (within-category) rankings by reversing the June 2002 change in the Morningstar rating methodology. Specifically, I rank funds within each category before June 2002 and rank all U.S. equity funds against each other staring in June 2002. Placebo distances to a rating threshold are calculated in an analogous way.

I combine Morningstar Direct data set with two standard mutual fund data sets: Center for Research in Security Prices (CRSP) Survivor-Bias-Free Mutual Fund Database and Thomson-Reuter Ownership Database (Thomson s12). I use CUSIPs to merge Morningstar and CRSP data sets and then use the MFLINKS files available through Wharton Research Data Services (WRDS) to merge CRSP and Thomson data sets. I obtain data on mutual fund monthly returns, total net assets (TNAs), and fund expenses from CRSP and the stock holdings of mutual funds are from Thomson. Although mutual fund share classes can receive different (within-category) percentile rankings and, as a result, can earn different star ratings, portfolio pumping activities would occur at the fund (portfolio) level. Therefore, for funds with multiple share classes, I aggregate share-class-level variables to the fund-level by computing the sum of total net assets, the maximum of fund ages, and the value-weighted average of other fund-level variables. I winsorize all variables at 1% and 99% each month. Table 2 shows the summary statistics on the fund-level variables.

[Insert Table 2]

3.2 Stock-Level Variables

In order to examine the threshold effect of star ratings on portfolio pumping at the level of mutual funds' stock holdings, I construct the stock-level variable $Ownership^{Near-Distant}$ by calculating the difference between the number of shares held by funds that are near a rating threshold (the distance below the bottom

20$^{\text{th}}$ percentile value) and the number of shares held by funds that are distant from rating thresholds (the distance above the top 20$^{\text{th}}$ percentile value), all scaled by the number of shares outstanding. I construct $Ownership^{Near-Distant}$ at the end of every calendar quarter from 1990Q1 to 2016Q3. Since Morningstar ratings are updated every month and portfolio pumping can occur at any month-end, all fund-level analyses are conducted at the fund-month level. For stock-level analyses, however, I focus on calendar quarter-ends in order to use the most accurate snapshot of stock holdings of mutual funds, following the literature on portfolio pumping (e.g., Carhart et al. (2002), Ben-David et al. (2019)).

The stock-level variable $Ownership^{Near-Distant}$ has a natural connection to the fund-level variable *Distance to a rating threshold*. To the extent that, compared to their distant peers, funds that are near a rating threshold are more like to engage in portfolio pumping, stocks that are held relatively more by funds that are near a rating threshold should be more likely to be pumped up. For placebo tests, I construct $Placebo\ ownership^{Near-Distant}$ in an analogous way, except that I use *placebo* percentile rankings that reverse the June 2002 change in the Morningstar rating methodology, as described in the previous section.

Next, I use quarterly estimates of two liquidity proxies that are widely used in the literature: Amihud (2002) illiquidity measure and Gibbs estimate of effective costs proposed by Hasbrouck (2004). Goyenko, Holden, and Trzcinka (2009) conduct horseraces of a large set of low-frequency liquidity measures calculated from the CRSP Daily Stock Files against high-frequency benchmarks calculated from the NYSE Trade and Quote (TAQ) data from 1993 to 2005 for a random sample of 400 stocks each year. And these authors find that Amihud's Illiquidity measure dominates in measuring price impact while Hasbrouck's Gibbs estimate of effective costs does very well in measuring effective and realized spreads. Hasbrouck (2009) also reaches the similar conclusion.

Amihud (2002) develops a price impact measure that captures the "daily price response associated with one million dollar of trading volume." Following Hasbrouck (2009), I use the squared root version to control for skewness in the original measure. Specifically, I use the ratio

$$Illiquidity = Average(\sqrt{\frac{|r_t|}{volume_t}}) \tag{1}$$

where where r_t is the stock return on day t and $volume_t$ is the dollar volume in million dollars on day t. The quarterly average is calculated over all positive days, since the ratio is undefined for zero-volume days.

Hasbrouck (2004) introduces a Gibbs sampler estimation of the half-spread ($c = \frac{1}{2}S$) in the Roll (1984) model using daily prices[8]. In the Roll model, the last observed trade price on day t is determined by

$$p_t = v_t + \frac{1}{2}Sq_t \tag{2}$$

where S is the effective spread, q_t is a buy/sell indicator for the last trade that equals +1 for a buy and −1 for a sell, and v_t is the unobserved fundamental value of the stock on day t, which evolves as

$$v_t = v_{t-1} + e_t \tag{3}$$

where e_t is the mean-zero, serially uncorrelated public information shock on day t.

Using the standard databases such as CRSP and Compustat, I further construct several stock characteristics including the logarithmic of market capitalization, the logarithmic of book-to-market ratio, the past 12-month cumulative stock return excluding the most recent month, the past one-month return excluding the last trading day of the month, and the number of analysts covering the stock from Institutional Brokers' Estimate System (I/B/E/S). All variables are measured as of the second-to-last trading day of the quarter. I winsorize all variables at 1% and 99% each quarter. To ensure that small infrequently traded stocks do not unduly influence my results, I remove micro-cap stocks using the NYSE 20th percentile breakpoint of market capitalization. Table 8 shows the summary statistics on the stock-level variables.

[Insert Table 8]

4 Fund-Level Evidence

In this section, I present fund-level evidence that discrete star ratings give fund managers powerful incentives to inflate their month-end performance when their funds are likely to finish the month in the

[8] Joel Hasbrouck generously provides the SAS code to compute the Gibbs estimator on his website.

vicinity of a rating threshold. Compared to their distant peers, mutual funds near a rating threshold earn abnormally higher returns on the last trading day of the month, which partially dissipates on the subsequent trading day. In the heterogeneity tests, the threshold effect of star ratings on net asset value (NAV) inflation is more pronounced among funds with grater incentives and abilities to pump up their portfolios. Placebo tests reversing the June 2002 change in the Morningstar rating methodology corroborate that the threshold effect of star ratings on portfolio pumping is likely causal.

4.1 Distance to a Rating Threshold and NAV Inflation

Given that portfolio pumping is an illegal practice, even if some managers pump up their portfolios, those managers are likely to do so when potential rewards are sufficiently large. That is, fund managers are likely to inflate month-end NAVs of their funds precisely when even a very small inflation of their fund performance, such as going from the 89th percentile to 90th percentile, can make a big difference, such as going from four stars to five stars. I test this prediction by exploiting the variation in the distance to a rating threshold. Since star ratings are updated every month and determined by fund performance over a rolling window, when a fund "rolls" out of a bad (good) month and/or "rolls" into a good (bad) month, the fund could move up (down) sharply in percentile rankings, generating considerable variation in percentile rankings and further variation in distances to rating thresholds (10^{th}, 32.5^{th}, 77.5^{th}, and 90^{th} percentiles).

Every month, I rank mutual funds on the basis of the Sharpe ratio calculated using daily returns in excess of the risk-free rate over the prior three years, but only through the second-to-last trading day of the month. Following Morningstar, I rank all U.S. equity funds against each other until May 2002 while ranking funds within each of Morningstar's 3×3 style categories starting in June 2002. Next, I calculate the distance to a rating threshold as the distance between a fund's Sharpe ratio percentile ranking and its nearest rating threshold on the second-to-last trading day of the month.

As a first path, I visually examine whether the distance to a rating threshold affects the month-end NAV inflation. To account for time fixed-effects, I subtract out the cross-sectional averages from fund returns on the last trading day of each month and on the first trading day of the subsequent month. Then, for each bin with the equal length of one-tenth of a percentile (i.e., 0.001), I average funds' adjusted-returns across each cross-section and calculate their time-series averages in the spirit of Fama and MacBeth (1973). In addition, I estimate the fitted line from the regression of the fund return on its squared distance to a rating threshold with a 95% confidence interval in the shaded area.

Figure 1 reports the results of the above analysis. Subfigure (a) shows that the distance to a rating threshold negatively predicts the fund return on the last trading day of the month. In addition, the effect of being closer to a rating threshold appears to be concave, suggesting that the incentive to inflate month-end performance diminishes more rapidly as funds are further away from a rating threshold. In contrast, Subfigure (b) shows that the distance to a rating threshold *positively* predicts the fund return on the first trading day of the subsequent month. This return reversal suggests that a large part of the gains earned on the last trading day of the month are transitory and driven by price impact, consistent with portfolio pumping.

[Insert Figure 1]

To formally examine the relation between the distance to a rating threshold and the month-end NAV inflation, I estimate the following linear regression model:

$$R_{i,t}^{Last\ day}\ (R_{i,t+1}^{First\ day}) = \beta \times Squared\ distance_{i,t} + \gamma \times Covariates_{i,t}\ (+\alpha_i) + \theta_t + \varepsilon_{i,t} \tag{4}$$

where i indexes mutual funds and t indexes time in month. The dependent variable is fund i's return on the last trading day of month t ($R_{i,t}^{Last\ day}$) or fund i's return on the first trading day of month $t+1$ ($R_{i,t+1}^{First\ day}$). The independent variable of interest, $Squared\ distance_{i,t}$, is fund i's squared distance between its Sharpe ratio percentile ranking and its nearest rating threshold on the *second-to-last* trading day of month t. $Covariates_{i,t}$ are a vector of fund characteristics that include the logarithmic of fund total net assets (TNA) (in million dollars), logarithmic of fund age (in years), expense ratio (in percent), turnover (in percent), an indicator variable for a load fund, and Sharpe ratio calculated using daily returns in excess of the risk-free rate over the prior three years, excluding the last trading day of month t. All independent variables are measured as of the *second-to-last* trading day of month t. Depending on the specification, the regression includes fund fixed-effects (α_i). All regressions include time fixed-effects (θ_t) and standard errors are double-clustered by fund and time.

The regression results are presented in Table 3. In the first three columns, the dependent variable is $R_{i,t}^{Last\ day}$. In column (1), the coefficent on $Squared\ distance_{i,t}$ is negative and statistically significant at the 1% level and remains virtually the same after the inclusion of a host of fund characteristics in column (2). This result is consistent with my prediction that funds near a rating threshold earn significantly higher returns on the last trading day of the month, compared to funds that are distant from all rating thresholds. The economic magnitude is large as well, given that the return is accumulated over just one day. Compared to funds that are about 10^{th} percentile away, funds near a rating threshold on average earn 144–167 basis points higher annualized returns on the last trading day of the month. Next, I exploit within-fund variation in the distance to a rating threshold by controlling for fund fixed-effects. In column (3), the coefficent on $Squared\ distance_{i,t}$ remains negative and statistically significant, although the economic magnitude is almost cut in half. Mutual funds tend to earn significantly higher returns on the last trading day of the month when they are more likely to finish in the vicinity of a rating threshold. In the last three columns, the dependent variable is $R_{i,t+1}^{First\ day}$. In all specifications, the coefficient on $R_{i,t+1}^{First\ day}$ is positive and statistically significant at conventional levels in baseline specifications with only time fixed-effects in columns (4) and (5). More than one third of the gains earned on the last trading day of the month dissipate on the subsequent trading day. This return reversal pattern around the turn-of-the-month corroborates that the gains on the last trading day of the month are transitory and likely driven by price impact, consistent with portfolio pumping.

[Insert Table 3]

4.2 Placebo Tests

In this subsection, I conduct placebo tests in order to provide corroborating evidence that the threshold effect of star ratings on NAV inflation documented in the previous subsection is likely causal. To this end, I exploit the June 2002 change in the Morningstar rating methodology. Morningstar ranked all U.S. equity funds against each other until May 2002, while ranking funds within each of Morningstar's 3×3 style categories starting in June 2002. I obtain *placebo* percentile rankings by reversing this June 2002 change in the Morningstar rating methodology. Specifically, placebo percentile rankings are based on each of Morningstar's finer categories until May 2002, whereas placebo percentile rankings are based on the broader category group starting in June 2002. Not surprisingly, placebo percentile rankings are highly correlated with actual percentile rankings (correlation coefficient = 0.80). Nevertheless, when measured in relation to rating thresholds (10th, 32.5th, 77.5th, and 90th percentiles), placebo distances are only weakly correlated with actual distances (correlation coefficient = 0.11).

With the placebo distance to a threshold, I estimate the following linear regression model:

$$R_{i,t}^{Last\ day}\ (R_{i,t+1}^{First\ day}) = \beta \times Squared\ placebo\ distance_{i,t} + \gamma \times Covariates_{i,t}\ (+\alpha_i) + \theta_t + \varepsilon_{i,t} \qquad (5)$$

where $Squared\ placebo\ distance_{i,t}$ is fund i's squared *placebo* distance between its Sharpe ratio *placebo* percentile ranking and its nearest rating threshold on the second-to-last trading day of month t, obtained by reversing the June 2002 change in the Morningstar rating methodology. The rest of the model is the same as Equation (4).

The regression results are presented in Table 4. The dependent variable is $R_{i,t}^{Last\ day}$ in columns (1) through (3). In column (1), the coefficient on $Squared\ placebo\ distance_{i,t}$ is negative, but small in magnitude and statistically insignificant. When I control for the fund characteristics and fund fixed-effects in columns (2) and (3), the coefficient on $Squared\ placebo\ distance_{i,t}$ even flips signs and remains statistically insignificant. In columns (4) through (6), I replace $R_{i,t}^{Last\ day}$ with $R_{i,t+1}^{First\ day}$ as the dependent variable and find that he coefficient on $Squared\ placebo\ distance_{i,t}$ is all negative (rather than positive) and insignificant. Thus, the *placebo* distance to a rating threshold does not have any meaningful impact on NAV inflation at the month-end. The null results in this subsection, combined with the significant results in the previous subsection, add further credibility to the claim that the threshold effect of star ratings on NAV inflation at the month-end is likely causal.

[Insert Table 4]

4.3 Heterogeneity Tests

4.3.1 Year- and Quarter-Ends vs. Other Month-Ends

In this subsection, I examine how the negative relation between the (squared) distance to a rating threshold and month-end NAV inflation varies across calendar months. Building on the convex flow–performance relation (Ippolito (1992), Sirri and Tufano (1998)), Carhart et al. (2002) propose the "leaning-for-the-tape" hypothesis as the main motivation for widespread portfolio pumping. While the leaning-for-the-tape effect implies the strongest portfolio pumping activities at the year-ends, the desire to obtain better star ratings does not preclude portfolio pumping at other month-ends. In order to corroborate that discrete star ratings incentivize some fund managers to pump, I examine whether fund managers pump up their portfolios throughout the course of a year. To this end, I re-estimate the baseline specification in Equation (4) in the sub-samples of year-ends, other quarter-ends, and other month-ends:

$$R_{i,t}^{Last\ day} = \beta \times Squared\ distance_{i,t} + \gamma \times Covariates_{i,t} + \theta_t + \varepsilon_{i,t} \tag{6}$$

The regression results are presented in Table 5. First, I examine portfolio pumping induced by star ratings at the year-ends. In the first two columns, the coefficient on $Squared\ distance_{i,t}$ is negative and statistically significant at conventional levels. The magnitude is much larger than that from the full sample in Section 4.1, suggesting that fund managers have a stronger incentive to finish the *year* with better star ratings. Next, I turn to the quarter-ends that are not year-ends. In the next two columns, the coefficient on $Squared\ distance_{i,t}$ is again negative and statistically significant at conventional levels, although the magnitude is much smaller than that for the year-ends. Last, I examine the month-ends that are not quarter-ends when the leaning-for-the-tape effect is likely to be minimal. In the last two columns, the coefficient on $Squared\ distance_{i,t}$ is still negative and statistically significant. Overall, these results suggest that mutual fund managers indeed care about star ratings, which induce some fund managers to inflate their month-end NAV when a small difference in performance rankings can make a discrete change in star ratings.

[Insert Table 5]

4.3.2 Young and Small Funds vs. Old and Large Funds

In this subsection, I examine whether the negative relation between the (squared) distance to a rating threshold and month-end NAV inflation is more pronounced among funds with a greater incentive to pump up their portfolios. Less established funds tend to have a greater flow–performance sensitivity, whereas investors flows for more established funds are less responsive to performance (Sirri and Tufano (1998)). As a result, fund managers with younger and smaller funds may be under a greater pressure to obtain better star ratings. In order to test this cross-sectional prediction, I add the interaction term with proxies for flow–performance sensitivity and estimate the following linear regression model:

$$\begin{aligned} R_{i,t}^{Last\ day} &= \delta \times Squared\ distance_{i,t} \times Sensitivity_{i,t} \\ &+ \beta \times Squared\ distance_{,i,t} + \rho \times Sensitivity_{i,t} + \gamma \times Covariates_{i,t}\ (+\alpha_t) + \theta_t + \varepsilon_{i,t} \end{aligned} \tag{7}$$

where i indexes mutual funds and t indexes time in month. The dependent variable is fund i's return on the last trading day of month t. For proxies for flow–performance sensitivity, $Sensitivity_{i,t}$, I use the logarithmic of fund age (in years) and logarithmic of fund total net assets (TNA) (in \$ million). The rest of the model is the same as Equation (4). Depending on the specification, the regression includes fund fixed-effects (α_i). All regressions include time fixed-effects (θ_t) and standard errors are double-clustered by fund and time.

The regression results are presented in Table 6. In the first three columns, the sensitivity proxy is the logarithmic of fund age. In the baseline specifications with only time fixed-effects in columns (1) and (2), the coefficient on $Squared\ distance_{i,t} \times \log(Fund\ age)_{i,t}$ is positive and statistically significant at conventional levels, with or without controls, implying that the threshold effect of star ratings on NAV inflation is larger among younger funds. When I control for fund fixed-effects, the results remain, albeit statically less significant, qualitatively similar. In the last three columns, the sensitivity proxy is replaced with the logarithmic of fund TNA. In all specifications, the coefficient on the interaction term is positive, as predicted. Although statistically less significant in the most parsimonious specification in column (4), the coefficient on $Squared\ distance_{i,t} \times \log(TNA)_{i,t}$ becomes statistically significant at the 5% level when I control for fund characteristics in column (5) and the results remain robust to further controlling for fund fixed-effects in column (6). Overall, these results suggest that the threshold effect of star ratings on month-end NAV inflation is more pronounced among funds with a greater incentive to pump up their portfolios, consistent with portfolio pumping activities driving this effect.

[Insert Table 6]

4.3.3 Small- and Mid-Cap Funds vs. Large-Cap Funds

In this subsection, I examine whether the negative relation between the (squared) distance to a rating threshold and month-end NAV inflation is more pronounced among funds with a greater ability to pump up their portfolios. If portfolio pumping is indeed responsible for the end-of-the-month NAV inflation, the effect on NAV inflation should be more pronounced among funds investing in less liquid stocks, since the closing prices of less liquid stocks would presumably be easier to influence (Carhart et al. (2002)). In order to test this prediction, I split funds into two groups by size category. To balance the number of fund-month observations between two sub-samples, I group small- and mid-cap funds together. To the extent that funds with a greatest ability to influence their NAVs are more likely to pump, the threshold effect of star ratings on NAV inflation at the month-end should be stronger among small- and mid-cap funds. In order to test this cross-sectional prediction, I re-estimate the linear regression model in Equation (4) in the sub-samples of funds split by size category:

$$R_{i,t}^{Last\ day} = \beta \times Squared\ distance_{i,t} + \gamma \times Covariates_{i,t}\ (+\alpha_t) + \theta_t + \varepsilon_{i,t} \tag{8}$$

The regression results are presented in Table 7. The results in columns (1) and (2) show that among small- and mid-cap funds, the squared distance to a rating threshold is negatively associated with the return on the last trading day of the month and the negative association is statistically significant at the 1% level before and after the inclusion of fund characteristics. In contrast, the negative relation between the squared distance to a rating threshold and the return on the last trading day of the month is much weaker among large-cap funds, as can be seen in columns (4) and (5). Adding fund fixed-effects to the baseline specification in columns (3) and (6) does not alter the conclusion that the threshold effect of star ratings on NAV inflation is stronger among funds investing in less liquid stocks, consistent with portfolio pumping driving this effect.

[Insert Table 7]

5 Holding-Level Evidence

In this section, I provide more direct evidence that the threshold effect of star ratings on NAV inflation is indeed driven by portfolio pumping. First, using mutual fund holdings data, I show that stocks predominantly held by funds that are near a rating threshold earn substantially higher returns on the last trading day of the month, compared to stocks predominantly held by funds that are distant from rating thresholds. By reversing the June 2002 change in the Morningstar rating methodology, I conduct placebo tests to mitigate concerns about endogenous mutual fund ownership and establish causality. Furthermore, the threshold effect of star ratings on month-end price inflation is more pronounced among less liquid stocks. Last, using NYSE's intraday Trade and Quote (TAQ) data, I show that the higher return on the last trading day of the month is concentrated in the last minutes of the trading session, consistent with portfolio pumping.

5.1 Evidence from Stock Holdings

In this subsection, I provide more direct evidence that the threshold effect of star ratings on NAV inflation is driven by portfolio pumping by examining mutual fund stock holdings. Specifically, I examine whether stock holdings of mutual funds that are close to a rating threshold earn significantly higher returns on the last trading day of the month, compared to stock holdings of mutual funds that are distant from rating thresholds. Although Morningstar ratings are updated every month and portfolio pumping can occur at any month-end, I focus on calendar quarter-ends in order to use the most accurate snapshot of stock holdings of mutual funds. Each quarter, I first rank all mutual funds based on its distance to a rating threshold, as defined in the previous section, and then define a fund to be "near" a rating threshold if the distance is below the bottom 20$^{\text{th}}$ percentile value and a fund to be "distant" from rating thresholds if the distance is above the top 20$^{\text{th}}$ percentile value. To measure the extent to which a stock is subject to portfolio pumping induced by rating thresholds, I calculate for each stock the ownership by funds that are near a rating threshold, subtracted by the ownership by funds that are distant from rating thresholds. Then, I estimate the following linear regression model:

$$R_{i,t}^{Last\ day}\ (R_{i,t+1}^{First\ day}) = \beta \times Ownership_{i,t}^{Near-Distant} + \gamma \times Covariates_{i,t}\ (+\alpha_t) + \theta_t + \varepsilon_{i,t} \tag{9}$$

where i indexes stocks and t indexes time in quarter. The dependent variable is stock i's return on the last trading day of quarter t ($R_{i,t}^{Last\ day}$) or stock i's return on the first trading day of quarter $t+1$ ($R_{i,t+1}^{First\ day}$). The independent variable of interest, $Ownership_{i,t}^{Near-Distant}$, is calculated as the difference between the number of shares held by funds that are near (the distance to a rating threshold below the bottom 20$^{\text{th}}$ percentile value) a rating threshold and the number of shares held by funds that are distant (the distance to a rating threshold above the top 20$^{\text{th}}$ percentile value) from rating thresholds, all scaled by the number of shares outstanding. $Covariates_{i,t}$ are a vector of stock characteristics that include the logarithmic of market capitalization, logarithmic of book-to-market ratio, past 12-month cumulative stock return excluding the most recent month, past one-month return excluding the last trading day of the month, number of analysts covering the stock, and quarterly estimates of Gibbs estimate of effective costs proposed by Hasbrouck (2004) and Amihud (2002) measure of price impact. Following Hasbrouck (2009), I use the squared root version of the Amihud (2002) illiquidity measure to control for skewness in the original measure. All independent variables are measured as of the second-to-last trading day of the quarter. Depending on the specification, the regression includes stock fixed-effects (α_i). All regressions include time fixed-effects (θ_t) and standard errors are double-clustered by stock and time.

The regression results are presented in Table 9. In column (1), the coefficient on $Ownership_{i,t}^{Near-Distant}$ is positive and statistically significant at the 1% level, suggesting that stocks that are predominantly held by funds near a rating threshold tend to earn significantly higher returns at the end of the quarter, compared to stocks that are predominantly held by funds distant from rating thresholds. This result is robust to the inclusion of a host of stock characteristics in columns (2). Next, I exploit within-stock variation in the ownership by funds near (or distant from) a rating threshold by controlling for stock fixed-effects. In column (3), the coefficient on $Ownership_{i,t}^{Near-Distant}$ remains positive and statistically significant, suggesting that stocks earn significantly higher returns at the end of the quarters when the stocks are predominantly held by funds near a rating threshold, compared to the end of the quarters when the same stocks are predominantly held by funds distant from rating thresholds. In the last three columns, the dependent variable is replaced with $R_{i,t+1}^{First\ day}$. In all specifications, the coefficient on $R_{i,t+1}^{First\ day}$ is negative and statistically significant at conventional levels. The comparison between the magnitude of coefficients suggests that virtually all gains earned on the last trading day of the quarter dissipate on the subsequent trading day. This return reversal pattern around the turn-of-the-month corroborates that the gains earned on the last trading day of the quarter are transitory and likely driven by price impact, consistent with portfolio pumping.

[Insert Table 9]

5.2 Robustness Checks and Placebo Tests

In my baseline results in the previous subsection, I have defined a fund to be "near" a rating threshold if the distance is below the bottom 20$^{\text{th}}$ percentile value and a fund to be "distant" from rating thresholds if the distance is above the top 20$^{\text{th}}$ percentile value. For robustness checks, I show that my results are robust to using different cutoffs to define "near" or "distant." I re-estimate the linear regression model in Equation (9) using 10$^{\text{th}}$ or 30$^{\text{th}}$ percentile values to define funds that are "near" or "distant" from a rating threshold, instead of the 20% percentile value used in the above results. The regression results are presented in Table 10. In Panel A, I use a tighter cutoff and define a fund to be "near" a rating threshold if its distance to a rating threshold is below the bottom 10$^{\text{th}}$ percentile value and a fund to be "distant" from rating thresholds if its distance to a rating threshold is above the top 10$^{\text{th}}$ percentile value. The results remain qualitatively similar and statistically significant at conventional levels. In Panel B, I use a looser cutoff and define a fund to be "near" a rating threshold if its distance to a rating threshold is below the bottom 30$^{\text{th}}$ percentile value and a fund to be "distant" from rating thresholds if its distance to a rating threshold is above the top 30$^{\text{th}}$ percentile value. The results remain, albeit statistically slightly less significant, qualitatively similar.

[Insert Table 10]

One concern with the results in the previous subsection is that mutual fund ownership is likely endogenous, and stock characteristics and stock fixed-effects may not adequately control for confounding factors. In order to address this concern and establish causality, I conduct placebo tests that are analogous to those in Section 4.2 by reversing the June 2002 change in the Morningstar rating methodology. Using the *placebo* distance to a rating threshold in Section 4.2, I define ownership by funds that are *placebo* near a rating threshold and ownership by funds that are *placebo* distant from all thresholds to construct relative *placebo* ownership. Then, I estimate the following linear regression model:

$$R_{i,t}^{Last\ day}\ (R_{i,t+1}^{First\ day}) = \beta \times Placebo\ ownership_{i,t}^{Near-Distant} + \gamma \times Covariates_{i,t}\ (+\alpha_t) + \theta_t + \varepsilon_{i,t} \tag{10}$$

where $Placebo\ ownership_{i,t}^{Near-Distant}$, is calculated as the difference between the number of shares held by funds that are *placebo* near (the *placebo* distance to a rating threshold below the bottom 20$^{\text{th}}$ percentile value) a rating threshold and the number of shares held by funds that are *placebo* distant (the *placebo* distance to a rating threshold above the top 20$^{\text{th}}$ percentile value) from rating thresholds, all scaled by the number of shares outstanding. The rest of the model is the same as Equation (9).

The regression results are presented in Table 11. The dependent variable is $R_{i,t}^{Last\ day}$ in the first three columns and $R_{i,t+1}^{First\ day}$ in the last three columns. In all specifications, the coefficient on $Placebo\ ownership_{i,t}^{Near-Distant}$ is small and statistically insignificant. These null results, combined with the significant results in the previous subsection, corroborate that the threshold effect of star ratings on price inflation is likely causal.

[Insert Table 11]

5.3 Heterogeneity Tests

In this subsection, I examine whether the threshold effect of star ratings on the end-of-the-month price inflation varies across different stock characteristics. If fund managers were to manipulate the closing prices of their stock holdings to inflate month-end NAVs of their funds, fund managers would focus on less liquid stocks, whose prices would presumably be much easier to influence. In order to test this cross-sectional prediction, I add the interaction term with measures of liquidity and estimate the following linear regression model:

$$\begin{aligned} R_{i,t}^{Last\ day} &= \delta \times Ownership_{i,t}^{Near-Distant} \times Liquidity_{i,t} \\ &\quad + \beta \times Ownership_{,i,t}^{Near-Distant} + \rho \times Liquidity_{i,t} + \gamma \times Covariates_{i,t}\ (+\alpha_t) + \theta_t + \varepsilon_{i,t} \end{aligned} \tag{11}$$

where i indexes stocks and t indexes time in quarter. The dependent variable is stock i's return on the last trading day of quarter t ($R_{i,t}^{Last\ day}$). For measures of liquidity, $Liquidity_{i,t}$, I use the Gibbs estimate of effective costs proposed by Hasbrouck (2004) and squared-root version of Amihud (2002) measure of price impact. The rest of the model is the same as Equation (9). Depending on the specification, the regression includes stock fixed-effects (α_i). All regressions include time fixed-effects (θ_t) and standard errors are double-clustered by stock and time.

The regression results are presented in Table 12. In the first three columns, the liquidity measure is the Gibbs estimate of effective costs. In the baseline specifications with only time fixed-effects in columns (1) and (2), the coefficient on $Ownership_{i,t}^{Near-Distant} \times Effective\ costs_{i,t}$ is positive and statistically significant, with or without controls, implying that less liquid stocks with larger spreads are more likely to be targeted for pumping. Next, I exploit within-stock variation in $Ownership_{i,t}^{Near-Distant}$ by controlling for stock fixed effects. In column (3), the coefficient of interest remains largely unchanged and statistically significant. In the last three columns, $Liquidity_{i,t}$ is replaced with Amihud (2002) measure of price impact. In the baseline specifications with only time fixed-effects in columns (4) and (5), the coefficient on $Ownership_{i,t}^{Near-Distant} \times$ $Price\ impact_{i,t}$ is positive and statistically significant, with or without controls, implying that less liquid stocks with larger price impact are more likely to be targeted for pumping. When I control for stock fixed-effects in column (6), the results remain, albeit statistically less significant, qualitatively similar. The heterogeneity tests at the holding-level in this subsection complement the heterogeneity tests at the fund-level in Section 4.3 that the effect of star ratings on month-end NAV inflation is much larger among small- and mid-cap funds. Overall, my results suggest that the threshold effect of star ratings on month-end price inflation is likely to be driven by portfolio pumping activities when funds are likely to finish the month in the vicinity of a rating threshold.

[Insert Table 12]

5.4 Intraday Returns

In this subsection, I provide more direct evidence on portfolio pumping induced by star ratings using NYSE's intraday Trade and Quote (TAQ) data. If the month-end inflation in NAVs of mutual funds near a rating threshold is indeed driven by fund managers' aggressive trading to influence closing prices of the stocks they own, the underlying stock price inflation is most likely to occur in the last minutes of the trading session. To test this prediction, I calculate intraday returns as log price changes over an interval of thirty minutes from 9:30 to 16:00 on the last trading day of the quarter and the first trading day of the subsequent quarter. Then, I estimate the following linear regression model:

$$R_{i,t}^{Last\ intraday} = \beta \times Ownership_{i,t}^{Near-Distant} + \gamma \times Covariates_{i,t} + \theta_t + \varepsilon_{i,t} \tag{12}$$

where i indexes stocks and t indexes time in quarter. The dependent variable is stock i's half-hour intraday return on the last trading day of quarter t ($R_{i,t}^{Last\ intraday}$). The rest of the model is the same as Equation (9). All regressions include time fixed-effects (θ_t) and standard errors are clustered by time.

The regression results are presented in Table 13. The coefficient on $Ownership_{i,t}^{Near-Distant}$ is small and statistically insignificant for all half-hour intraday returns from 9:30 to 15:30 in columns (1) through (12). However, the coefficient becomes large and statistically significant at the 1% level for the last half-hour intraday return from 15:30 to 16:00 in column (13). Consistent with portfolio pumping, the significantly higher returns earned by stocks predominantly held by funds near a rating threshold on the last trading day of the quarter are highly concentrated in the last minutes of the trading session. Overall, the intraday evidence further supports my claim that the threshold effect of star ratings on price inflation at the month-end is indeed driven by portfolio pumping activities.

[Insert Table 13]

6 Conclusion

The idea that mutual fund investors use risk-adjusted returns in their capital allocation decision is appealing and has motivated some prominent financial economists to use mutual fund flows to test asset pricing models. Berk and van Binsbergen (2016), for instance, conclude that the CAPM is the "closest to the asset pricing model investors are actually using". Using a different methodology, Barber, Huang, and Odean (2016) also reach the same conclusion. In a recent study, however, Ben-David et al. (2019) show that simple and readily available star ratings explain mutual fund investors' behavior much better than any asset pricing models. That is, star ratings are the main determinant of capital allocation across mutual funds, as first suggested by Del Guercio and Tkac (2008). Reuter and Zitzewitz (2015) estimate that about 22% of the difference in future fund flows received by five-star funds and one-star funds represents a causal effect of the difference in star ratings on investor flows. Given that mutual fund investors rely heavily on star ratings, it is natural to presume that fund managers could be pressured to maintain desirable star ratings.

This paper reveals that the discrete nature of Morningstar ratings, i.e., one star to five stars, gives mutual fund managers powerful incentives to inflate their month-end performance when their funds are likely to finish the month in the vicinity of a rating threshold. As their percentile rankings move closer to a rating threshold, mutual funds experience substantially larger gains on the last trading day of the month, which partially dissipate on the next trading day. In addition, stocks predominantly held by funds close to a rating cutoff also earn significantly higher returns at the month-end, especially during the last minutes of the trading session. Placebo tests exploiting a change in the Morningstar rating methodology around June 2002 provide corroborating evidence that the threshold effect of star ratings on portfolio pumping is likely causal.

References

Amihud, Yakov, 2002, Illiquidity and Stock Returns: Cross-Section and Time-Series Effects, *Journal of Financial Markets* 5, 31–56.

Barber, Brad M., Xing Huang, and Terrance Odean, 2016, Which Factors Matter to Investors? Evidence from Mutual Fund Flows, *Review of Financial Studies* 29, 2600–2642.

Ben-David, Itzhak, Francesco Franzoni, Augustin Landier, and Rabih Moussawi, 2013, Do Hedge Funds Manipulate Stock Prices?, *Journal of Finance* 68, 2383–2434.

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, 2019, What Do Mutual Fund Investors Really Care About?, *Working Paper, Ohio State University and NBER* .

Berk, Jonathan B., and Jules H. van Binsbergen, 2016, Assessing asset pricing models using revealed preference, *Journal of Financial Economics* 119, 1–23.

Carhart, Mark M, 1997, On Persistence in Mutual Fund Performance, *Journal of Finance* 52, 57–82.

Carhart, Mark M., Ron Kaniel, David K. Musto, and Adam V. Reed, 2002, Leaning for the Tape: Evidence of Gaming Behavior in Equity Mutual Funds, *Journal of Finance* 57, 661–693.

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

Del Guercio, Diane, and Paula A Tkac, 2008, Star Power: The Effect of Morningstar Ratings on Mutual Fund Flow, *Journal of Financial and Quantitative Analysis* 43, 907–936.

Goyenko, Ruslan Y., Craig W. Holden, and Charles A. Trzcinka, 2009, Do liquidity measures measure liquidity?, *Journal of Financial Economics* 92, 153–181.

Hasbrouck, Joel, 2004, Liquidity in the futures pits: Inferring market dynamics from incomplete data, *Journal of Financial and Quantitative Analysis* 39, 305–326.

Hasbrouck, Joel, 2009, Trading costs and returns for U. S. equities: Estimating effective costs from daily data, *Journal of Finance* 64, 1445–1477.

Hu, Gang, R David Mclean, Jeffrey Pontiff, and Qinghai Wang, 2014, The Year-End Trading Activities of Institutional Investors: Evidence from Daily, *Review of Financial Studies* 27, 1593–1614.

Ippolito, Richard A., 1992, Consumer Reaction to Measures of Poor Quality: Evidence from the Mutual Fund Industry, *Journal of Law and Economics* 35, 45–70.

Jensen, Michael C ., 1969, Risk, The Pricing of Capital Assets, and The Evaluation of Investment Portfolios, *Journal of Business* 42, 167–247.

Ni, Sophie Xiaoyan, Neil D. Pearson, and Allen M. Poteshman, 2005, Stock price clustering on option expiration dates, *Journal of Financial Economics* 78, 49–87.

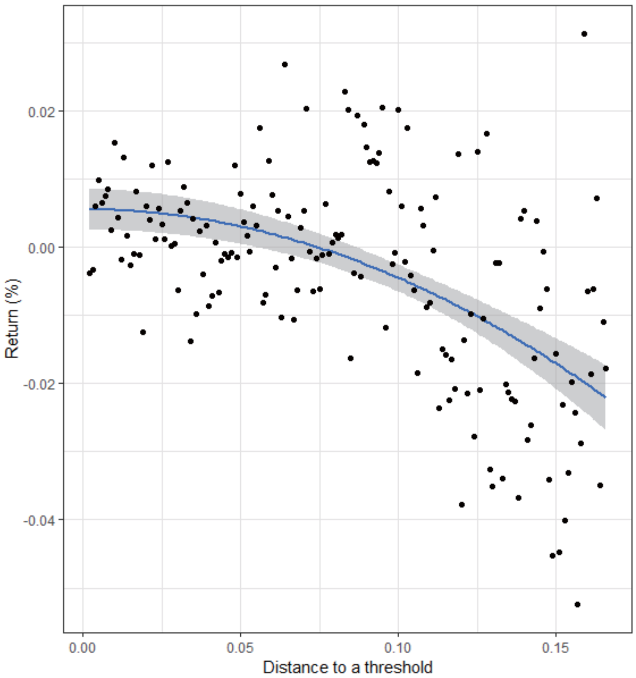
Reuter, Jonathan, and Eric Zitzewitz, 2015, How Much Does Size Erode Mutual Fund Performance? A Regression Discontinuity Approach, *Working Paper, Boston College and NBER* .

Roll, Richard, 1984, A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market, *Journal of Finance* 39, 1127–1139.

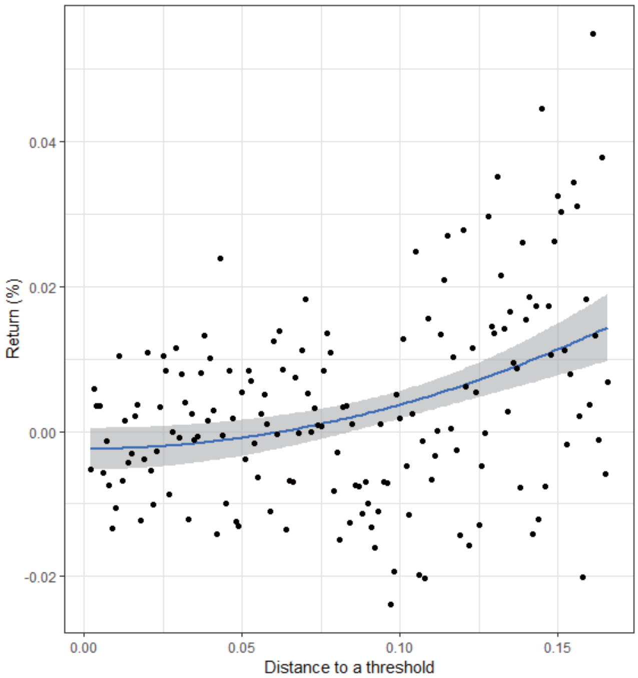
Sharpe, William F, 1998, Morningstar's Risk-Adjusted Ratings, *Financial Analysts Journal* 54, 21–33.

Sirri, Erik R, and Peter Tufano, 1998, Costly Search and Mutual Fund Flows, *Journal of Finance* 53, 1589–1622.

Zweig, Jason, 1997, Watch out for the year-end fund flimflam, *Money Magazine* November 1, 130–133.



(a) Last day of month t



(b) First day of month $t+1$

Figure 1: Distance to a Rating Threshold and NAV Inflation

This figure plots the average returns on the last trading day of month t in subfigure (a) and on the first trading day of month $t+1$ in subfigure (b) across equally spaced bins of the distance to a rating threshold on the second-to-last trading day of month t. First, I subtract out the cross-sectional average returns from fund returns to account for time fixed-effects. Then, for each bin with the equal length of one-tenth of a percentile (i.e., 0.001), I average funds' adjusted-returns across each cross-section and report their time-series averages in the spirit of Fama and MacBeth (1973). In addition, I report the fitted line from the regression of the fund return on its squared distance to a rating threshold with a 95% confidence interval in the shaded area.

Table 1: Transition Probabilities of Morningstar Ratings

This table reports the transition probabilities of Morningstar ratings. Mutual fund share classes are rated every month on a scale of one star (lowest) to five stars (highest) on the basis of Morningstar "Risk-Adjusted Returns" (RAR) over the prior three, five, and ten years, depending on data availability. On the basis of Morningstar RAR percentile rankings, the top 10% of funds receive five stars, the next 22.5% four stars, the middle 35% three stars, the next 22.5% two stars, and the bottom 10% one star. Morningstar ranked all U.S. equity funds against each other until May 2002, while raking funds within each of Morningstar's 3×3 (size interacted with value) categories starting in June 2002.

	Rating in month t				
Rating in month $t-1$	1	2	3	4	5
1	87.76	11.99	0.23	0.02	0.004
2	3.81	86.82	9.26	0.10	0.004
3	0.04	6.49	87.49	5.92	0.05
4	0.01	0.11	10.55	85.25	4.09
5	0.001	0.03	0.26	14.08	85.63

Table 2: Summary Statistics on Fund-Level Variables

This table reports the summary statistics on the fund-level variables. My primary data come from Morningstar Direct, from which I obtain data on daily returns, Morningstar categories, star ratings, inception dates, and index-fund indicator. Morningstar ratings are updated every month and based on Morningstar "Risk-Adjusted Return" (RAR), which adjusts for risk and accounts for sales charges using monthly returns. In order to obtain percentile rankings that are not contaminated by the outcomes of the last trading session of the month, while replicating percentile rankings based on Morningstar RAR as closely as possible, I calculate the Sharpe ratio of each mutual fund share class every month using daily returns in excess of the risk-free rate over the prior three years excluding the last trading day. There are four rating thresholds dividing all funds into five star ratings: 10^{th}, 32.5^{th}, 77.5^{th}, and 90^{th} percentiles. The distance to a rating threshold is calculated for each share class every month as the distance between its Sharpe ratio percentile ranking and its nearest threshold on the second-to-last trading day of the month. Following Morningstar, I rank all U.S. equity funds against each other until May 2002, while ranking funds within each of Morningstar's 3×3 categories starting in June 2002. Although I focus on on actively-managed funds, I keep index funds in the ranking procedure in line with the Morningstar rating methodology. To test placebo effects, I calculate *placebo* percentile rankings by reversing the June 2002 change in the Morningstar rating methodology. Specifically, I rank funds within each of Morningstar's 3×3 categories until May 2002, while ranking all U.S. equity funds against each other staring in June 2002. Placebo distances to a rating threshold are calculated in an analogous way. I combine Morningstar Direct data set with two standard mutual fund data sets: Center for Research in Security Prices (CRSP) Survivor-Bias-Free Mutual Fund Database and Thomson-Reuter Ownership Database (Thomson s12). I use CUSIPs to merge Morningstar and CRSP data sets and then use the MFLINKS files available through Wharton Research Data Services (WRDS) to merge CRSP and Thomson data sets. I obtain data on mutual fund monthly returns, total net assets (TNAs), and fund expenses from CRSP and the stock holdings of mutual funds are from Thomson. Although mutual fund share classes can receive different percentile rankings and, as a result, different star ratings, portfolio pumping would occur at the fund (portfolio) level. Therefore, for funds with multiple share classes, I aggregate share-class-level variables to the fund-level by computing the sum of total net assets, the maximum of fund ages, and the value-weighted average of other fund-level variables. I winsorize all variables at 1% and 99% each month. The sample covers the period from 1990 to 2018.

Variable	Obs.	Mean	St. Dev.	Q_1	Median	Q_3
$R_t^{\text{Last Day}}$ (%)	476,263	0.07	1.02	−0.47	0.03	0.57
Distance to a rating threshold	476,263	0.07	0.04	0.03	0.06	0.09
Placebo distance to a rating threshold	476,263	0.07	0.04	0.03	0.06	0.09
TNA ($million)	476,263	1,231.82	3,065.74	65.49	247.44	940.65
Fund age (in years)	476,263	15.16	12.73	6.66	11.41	18.76
Expense ratio (%)	476,263	1.21	0.49	0.93	1.15	1.42
Turnover (%)	476,263	82.60	86.74	32	60	103
$\mathbb{1}$(Load fund)	476,263	0.57	0.49	0	1	1
Sharpe ratio	476,263	0.03	0.03	0.01	0.04	0.06

Table 3: Distance to a Rating Threshold and NAV Inflation

This table presents the results of the following linear regression model:

$$R_{i,t}^{Last\ day}\ (R_{i,t+1}^{First\ day}) = \beta \times Squared\ distance_{i,t} + \gamma \times Covariates_{i,t}\ (+\alpha_i) + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time in month. The dependent variable is fund i's return on the last trading day of month t in columns (1) through (3) and fund i's return on the first trading day of month $t+1$ in columns (4) through (6). The independent variable of interest, *Squared distance*$_{i,t}$, is fund i's squared distance between its Sharpe ratio percentile ranking and its nearest rating threshold on the *second-to-last* trading day of month t. More details on the construction of *Distance to a rating threshold* are provided in Table 2. *Covariates*$_{i,t}$ are a vector of fund characteristics that include the logarithmic of fund total net assets (TNA) (in $ million), logarithmic of fund age (in years), expense ratio (in percent), turnover (in percent), an indicator variable for a load fund, and Sharpe ratio calculated using fund daily returns in excess of the risk-free rate over the prior three years, excluding the last trading day of month t. All independent variables are measured as of the *second-to-last* trading day of month t. In columns (3) and (6), the regression includes fund fixed-effects (α_i). All regressions include time fixed-effects (θ_t), standard errors are double-clustered by fund and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. The sample covers the period from 1990 to 2018.

Dependent variable:	$R_t^{\text{Last Day}}$ (%)			$R_{i,t+1}^{First\ day}$ (%)		
	(1)	(2)	(3)	(4)	(5)	(6)
Squared distance	-0.66^{***}	-0.57^{***}	-0.36^{***}	0.26^{*}	0.24^{*}	0.16
	(−3.93)	(−3.38)	(−2.65)	(1.73)	(1.67)	(1.17)
log(TNA)		0.001	−0.01		−0.001	−0.004
		(0.51)	(−1.53)		(−0.59)	(−0.94)
log(Fund age)		-0.01^{***}	-0.03^{***}		0.01^{**}	0.02^{*}
		(−3.37)	(−3.13)		(2.54)	(1.75)
Expense ratio (%)		0.04^{***}	0.02^{**}		-0.02^{***}	−0.01
		(5.22)	(2.55)		(−3.81)	(−1.03)
Turnover (%)		0.0001^{*}	0.0001^{**}		0.0000	−0.0000
		(1.88)	(2.03)		(0.37)	(−0.08)
$\mathbb{1}$(Load fund)		0.002	−0.001		0.003	0.004
		(0.53)	(−0.30)		(1.35)	(1.24)
Sharpe ratio		1.87^{***}	1.62^{***}		1.18^{**}	1.43^{**}
		(3.84)	(2.81)		(2.08)	(2.21)
Time fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Fund fixed-effects	No	No	Yes	No	No	Yes
Observations	476,263	476,263	476,263	476,263	476,263	476,263
Adjusted R^2	0.81	0.81	0.82	0.87	0.87	0.87

Table 4: Distance to a Rating Threshold and NAV Inflation: Placebo Tests

This table presents the results of the following linear regression model:

$$R_{i,t}^{Last\ day}\ (R_{i,t+1}^{First\ day}) = \beta \times Squared\ placebo\ distance_{i,t} + \gamma \times Covariates_{i,t}\ (+\alpha_i) + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time in month. The dependent variable is fund i's return on the last trading day of month t in columns (1) through (3) and fund i's return on the first trading day of month $t+1$ in columns (4) through (6). The independent variable of interest, $Squared\ placebo\ distance_{i,t}$, is fund i's squared distance between its Sharpe ratio *placebo* percentile ranking and its nearest rating threshold on the *second-to-last* trading day of month t. I obtain *placebo* percentile rankings by reversing the June 2002 change in the Morningstar rating methodology. That is, I rank funds within each of Morningstar's 3×3 categories until May 2002, while ranking all U.S. equity funds against each other starting in June 2002. $Covariates_{i,t}$ are a vector of fund characteristics that include the logarithmic of fund total net assets (TNA) (in \$ million), logarithmic of fund age (in years), expense ratio (in percent), turnover (in percent), an indicator variable for a load fund, and Sharpe ratio calculated using fund daily returns in excess of the risk-free rate over the prior three years, excluding the last trading day of month t. All independent variables are measured as of the *second-to-last* trading day of month t. In columns (3) and (6), the regression includes fund fixed-effects (α_i). All regressions include time fixed-effects (θ_t), standard errors are double-clustered by fund and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. The sample covers the period from 1990 to 2018.

Dependent variable:	$R_t^{\text{Last Day}}$ (%)			$R_{t+1}^{\text{First Day}}$ (%)		
	(1)	(2)	(3)	(4)	(5)	(6)
Squared *placebo* distance	−0.19	0.001	0.09	−0.12	−0.09	−0.17
	(−1.26)	(0.01)	(0.68)	(−0.77)	(−0.59)	(−1.12)
log(TNA)		0.001	−0.01		−0.001	−0.004
		(0.52)	(−1.54)		(−0.59)	(−0.94)
log(Fund age)		−0.01***	−0.03***		0.01**	0.02*
		(−3.37)	(−3.12)		(2.54)	(1.75)
Expense ratio (%)		0.04***	0.02**		−0.02***	−0.01
		(5.25)	(2.56)		(−3.82)	(−1.04)
Turnover (%)		0.0001*	0.0001**		0.0000	−0.0000
		(1.87)	(2.03)		(0.37)	(−0.08)
$\mathbb{1}$(Load fund)		0.002	−0.001		0.003	0.004
		(0.57)	(−0.30)		(1.31)	(1.23)
Sharpe ratio		1.87***	1.63***		1.18**	1.43**
		(3.84)	(2.81)		(2.08)	(2.20)
Time fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Fund fixed-effects	No	No	Yes	No	No	Yes
Observations	476,263	476,263	476,263	476,263	476,263	476,263
Adjusted R^2	0.81	0.81	0.82	0.87	0.87	0.87

Table 5: Year- and Quarter-Ends vs. Other Month-Ends

This table presents the results of the following linear regression model:

$$R_{i,t}^{Last\ day} = \beta \times Squared\ distance_{i,t} + \gamma \times Covariates_{i,t} + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time in month. The results for year-ends are reported in columns (1) and (2), for quarter-ends that are not year-ends in columns (3) and (4), and for month-ends that are not quarter-ends in columns (5) and (6). The dependent variable is fund i's return on the last trading day of month t. The independent variable of interest, $Squared\ distance_{i,t}$, is fund i's squared distance between its Sharpe ratio percentile ranking and its nearest rating threshold on the *second-to-last* trading day of month t. $Covariates_{i,t}$ are a vector of fund characteristics that include the logarithmic of fund total net assets (TNA) (in $ million), logarithmic of fund age (in years), expense ratio (in percent), turnover (in percent), an indicator variable for a load fund, and Sharpe ratio calculated using fund daily returns in excess of the risk-free rate over the prior three years, excluding the last trading day of month t. All independent variables are measured as of the *second-to-last* trading day of month t. All regressions include time fixed-effects (θ_t), standard errors are double-clustered by fund and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. The sample covers the period from 1990 to 2018.

Dependent variable:	$R_t^{\text{Last Day}}$ (%)					
	Year-Ends		Other Quarter-Ends		Other Month-Ends	
	(1)	(2)	(3)	(4)	(5)	(6)
Squared distance	-1.49^{**}	-1.41^{*}	-0.73^{**}	-0.60^{*}	-0.52^{***}	-0.44^{**}
	(-2.00)	(-1.89)	(-2.07)	(-1.65)	(-2.80)	(-2.42)
log(TNA)		0.004		-0.001		0.001
		(0.77)		(-0.59)		(0.73)
log(Fund age)		-0.03^{**}		-0.02^{***}		-0.01^{**}
		(-2.28)		(-2.59)		(-1.99)
Expense ratio (%)		0.06^{***}		0.05^{***}		0.03^{***}
		(2.91)		(4.21)		(3.47)
Turnover (%)		-0.0001		0.0001^{**}		0.0001
		(-0.80)		(1.98)		(1.51)
$\mathbb{1}$(Load fund)		-0.003		0.0003		0.003
		(-0.48)		(0.06)		(1.12)
SR_3yr		-1.99		3.21^{***}		1.87^{***}
		(-1.16)		(2.84)		(3.52)
Time fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	40,865	40,865	119,640	119,640	315,758	315,758
Adjusted R^2	0.77	0.78	0.81	0.82	0.81	0.82

Table 6: Young and Small Funds vs. Old and Large Funds

This table presents the results of the following linear regression models:

$$R_{i,t}^{Last\ day} = \delta \times Squared\ distance_{i,t} \times Sensitivity_{i,t} + \beta \times Squared\ distance_{,i,t} + \rho \times Sensitivity_{i,t} + \gamma \times Covariates_{i,t}\ (+\alpha_t) + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time in month. The dependent variable is fund i's return on the last trading day of month t. For proxies for flow–performance sensitivity, $Sensitivity_{i,t}$, I use the logarithmic of fund age (in years) and logarithmic of fund total net assets (TNA) (in $ million). $Squared\ distance_{i,t}$ is fund i's squared distance between its Sharpe ratio percentile ranking and its nearest rating threshold on the *second-to-last* trading day of month t. $Covariates_{i,t}$ are a vector of fund characteristics that include the logarithmic of fund total net assets (TNA) (in $ million), logarithmic of fund age (in years), expense ratio (in percent), turnover (in percent), an indicator variable for a load fund, and Sharpe ratio calculated using fund daily returns in excess of the risk-free rate over the prior three years, excluding the last trading day of month t. All independent variables are measured as of the *second-to-last* trading day of month t. In columns (3) and (6), the regression includes fund fixed-effects (α_i). All regressions include time fixed-effects (θ_t), standard errors are double-clustered by fund and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. The sample covers the period from 1990 to 2018.

Dependent variable:	$R_t^{\text{Last Day}}$ (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
Sq. distance × log(Fund age)	0.35**	0.33**	0.17			
	(2.24)	(2.08)	(1.13)			
Sq. distance × log(TNA)				0.07	0.13**	0.15**
				(1.02)	(2.13)	(2.56)
Squared distance	−1.52***	−1.37***	−0.79*	−1.04**	−1.30***	−1.20***
	(−3.15)	(−2.84)	(−1.81)	(−2.38)	(−3.15)	(−3.21)
log(Fund age)	−0.02***	−0.01***	−0.03***		−0.01***	−0.03***
	(−4.71)	(−3.61)	(−3.20)		(−3.37)	(−3.13)
log(TNA)		0.001	−0.01	−0.002	−0.0001	−0.01*
		(0.51)	(−1.53)	(−1.15)	(−0.09)	(−1.71)
Expense ratio (%)		0.04***	0.02**		0.04***	0.02**
		(5.22)	(2.55)		(5.22)	(2.54)
Turnover (%)		0.0001*	0.0001**		0.0001*	0.0001**
		(1.88)	(2.03)		(1.88)	(2.04)
$\mathbb{1}$(Load fund)		0.002	−0.001		0.002	−0.001
		(0.53)	(−0.30)		(0.53)	(−0.30)
Sharpe ratio		1.87***	1.62***		1.87***	1.63***
		(3.84)	(2.81)		(3.85)	(2.82)
Time fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Fund fixed-effects	No	No	Yes	No	No	Yes
Observations	476,263	476,263	476,263	476,263	476,263	476,263
Adjusted R^2	0.81	0.81	0.82	0.81	0.81	0.82

Table 7: Small- and Mid-Cap Funds vs. Large-Cap Funds

This table presents the results of the following linear regression model:

$$R_{i,t}^{Last\ day} = \beta \times Squared\ distance_{i,t} + \gamma \times Covariates_{i,t}\ (+\alpha_i) + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time in month. I split funds into two groups by size category. To balance the number of fund-month observations between two sub-samples, I group small- and mid-cap funds together. I report the results for small- and mid-cap funds in columns (1) through (3) and the results for large-cap funds in columns (4) through (6). The dependent variable is fund i's return on the last trading day of month t. The independent variable of interest, $Squared\ distance_{i,t}$, is fund i's squared distance between its Sharpe ratio percentile ranking and its nearest rating threshold on the *second-to-last* trading day of month t. $Covariates_{i,t}$ are a vector of fund characteristics that include the logarithmic of fund total net assets (TNA) (in $ million), logarithmic of fund age (in years), expense ratio (in percent), turnover (in percent), an indicator variable for a load fund, and Sharpe ratio calculated using fund daily returns in excess of the risk-free rate over the prior three years, excluding the last trading day of month t. All independent variables are measured as of the *second-to-last* trading day of month t. In columns (3) and (6), the regression includes fund fixed-effects (α_i). All regressions include time fixed-effects (θ_t), standard errors are double-clustered by fund and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. The sample covers the period from 1990 to 2018.

Dependent variable:	$R_{i,t}^{Last\ day}$					
	Small- and mid-cap funds			Large-cap funds		
	(1)	(2)	(3)	(4)	(5)	(6)
Squared distance	−0.63***	−0.57***	−0.47**	−0.23	−0.14	−0.06
	(−3.08)	(−2.78)	(−2.43)	(−1.51)	(−0.96)	(−0.39)
log(TNA)		0.004**	−0.004		−0.001	−0.004
		(1.99)	(−0.60)		(−0.89)	(−1.19)
log(Fund age)		−0.01**	−0.03**		0.001	0.005
		(−2.20)	(−2.33)		(0.68)	(0.75)
Expense ratio (%)		0.02***	0.02		0.01**	0.01*
		(3.33)	(1.48)		(2.35)	(1.91)
Turnover (%)		0.0001*	0.0001		0.0000	0.0000
		(1.79)	(1.23)		(1.04)	(1.40)
$\mathbb{1}$(Load fund)		0.01**	0.003		0.002	0.003
		(2.08)	(0.45)		(0.89)	(0.79)
Sharpe ratio		1.37***	1.66***		2.30***	2.30***
		(3.21)	(2.91)		(4.60)	(3.52)
Time fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Fund fixed-effects	No	No	Yes	No	No	Yes
Observations	217,716	217,716	217,716	258,547	258,547	258,547
Adjusted R^2	0.82	0.82	0.83	0.88	0.88	0.88

Table 8: Summary Statistics on Stock-Level Variables

This table reports the summary statistics on the stock-level variables. I construct $Ownership^{Near-Distant}$ by calculating the difference between the number of shares held by funds that are near a rating threshold (the distance below the bottom 20th percentile value) and the number of shares held by funds that are distant from rating thresholds (the distance above the top 20th percentile value), all scaled by the number of shares outstanding. To test placebo effects, I construct *Placebo* $ownership^{Near-Distant}$ in an analogous way, except that I use the *placebo* distance to a rating threshold as described in Table 2. I use quarterly estimates of two liquidity proxies: Gibbs estimate of effective costs proposed by Hasbrouck (2004) and price impact measure of Amihud (2002). Following Hasbrouck (2009), I use the squared root version of the Amihud (2002) illiquidity measure to control for skewness in the original measure. Other stock characteristics include the logarithmic of market capitalization, logarithmic of book-to-market ratio, past 12-month cumulative stock return excluding the most recent month, past one-month return excluding the last trading day of the month, and number of analysts covering the stock, all measured as of the second-to-last trading day of the quarter. I winsorize all variables at 1% and 99% each quarter. I also remove all micro-cap stocks using the NYSE 20th percentile breakpoint of market capitalization in the previous month. The sample covers the period from 1990 to 2017.

Variable	Obs.	Mean	St. Dev.	Q_1	Median	Q_3
$\text{Ownership}^{\text{Near-Distant}}$	236,692	−0.03	0.03	−0.04	−0.02	−0.003
Placebo $\text{ownership}^{\text{Near-Distant}}$	236,692	−0.02	0.03	−0.03	−0.01	−0.000
$R_t^{\text{Last Day}}$ (%)	236,692	0.35	3.08	−1.08	0.00	1.44
Market cap (in $million)	236,692	4,043.93	10,312.79	354.03	857.43	2,698.94
Book-to-market	236,692	0.58	0.47	0.28	0.48	0.75
One-month return	236,692	0.02	0.12	−0.04	0.01	0.07
Twelve-month return	236,692	0.21	0.62	−0.10	0.11	0.37
Analyst Coverage	236,692	8.75	7.00	4	7	12
Gibbs effective costs	236,692	0.59	0.42	0.31	0.47	0.74
Amihud's price impact	236,692	0.12	0.19	0.03	0.06	0.14

Table 9: Evidence from Stock Holdings

This table presents the results of the following linear regression model:

$$R_{i,t}^{Last\ day}\ (R_{i,t+1}^{First\ day}) = \beta \times Ownership_{i,t}^{Near-Distant} + \gamma \times Covariates_{i,t}\ (+\alpha_t) + \theta_t + \varepsilon_{i,t}$$

where i indexes stocks and t indexes time in quarter. The dependent variable is stock i's return on the last trading day of quarter t in columns (1) through (3) and stock i's return on the first trading day of quarter $t+1$ in columns (4) through (6). The independent variable of interest, $Ownership_{i,t}^{Near-Distant}$, is calculated as the difference between the number of shares held by funds that are near a rating threshold (the distance below the bottom 20$^{\text{th}}$ percentile value) and the number of shares held by funds that are distant from rating thresholds (the distance above the top 20$^{\text{th}}$ percentile value), all scaled by the number of shares outstanding. $Covariates_{i,t}$ are a vector of stock characteristics that include the logarithmic of market capitalization, logarithmic of book-to-market ratio, past 12-month cumulative stock return excluding the most recent month, past one-month return excluding the last trading day of the month, number of analysts covering the stock, and quarterly estimates of Gibbs estimate of effective costs proposed by Hasbrouck (2004) and squared-root version of Amihud (2002) measure of price impact. All independent variables are measured as of the second-to-last trading day of the quarter. In columns (3) and (6), the regression includes stock fixed-effects (α_i). All regressions include time fixed-effects (θ_t), standard errors are double-clustered by stock and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. The sample covers the period from 1990 to 2017.

Dependent variable:	$R_t^{\text{Last Day}}$ (%)			$R_{t+1}^{\text{First Day}}$ (%)		
	(1)	(2)	(3)	(4)	(5)	(6)
Ownership$^{\text{Near-Distant}}$	1.04***	0.96***	0.78**	−1.22***	−1.00***	−0.77**
	(2.99)	(2.82)	(2.36)	(−3.07)	(−2.70)	(−2.14)
log(Market cap)		−0.09***	−0.09		0.05**	−0.27***
		(−3.34)	(−1.39)		(2.05)	(−3.68)
log(Book-to-market)		−0.04	−0.003		0.17***	0.10*
		(−1.33)	(−0.07)		(4.35)	(1.79)
One-month return		−1.40***	−1.39***		−0.27	−0.03
		(−4.91)	(−5.17)		(−0.62)	(−0.07)
Twelve-month return		0.13	0.08		−0.26	−0.17
		(1.56)	(0.97)		(−1.20)	(−0.85)
Analyst coverage		−0.002	−0.12**		−0.02	0.08
		(−0.07)	(−2.27)		(−0.84)	(1.52)
Gibbs' effective costs		0.40***	0.24*		−0.17	−0.08
		(2.94)	(1.88)		(−0.94)	(−0.59)
Amihud's price impact		−0.21	0.32		0.26	0.14
		(−0.74)	(1.07)		(0.66)	(0.38)
Time fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Stock fixed-effects	No	No	Yes	No	No	Yes
Observations	236,692	236,692	236,692	236,692	236,692	236,692
Adjusted R^2	0.14	0.14	0.17	0.21	0.21	0.23

Table 10: Evidence from Stock Holdings: Robustness Checks

This table presents the results of the following linear regression model:

$$R_{i,t}^{Last\ day}\ (R_{i,t+1}^{First\ day}) = \beta \times Ownership_{i,t}^{Near-Distant} + \gamma \times Covariates_{i,t}\ (+\alpha_t) + \theta_t + \varepsilon_{i,t}$$

where i indexes stocks and t indexes time in quarter. The dependent variable is stock i's return on the last trading day of quarter t in columns (1) through (3) and stock i's return on the first trading day of quarter $t+1$ in columns (4) through (6). In Panel A (B), $Ownership_{i,t}^{Near-Distant}$ is calculated as the difference between the number of shares held by funds that are near a rating threshold (the distance below the bottom 10^{th} (30^{th}) percentile value) and the number of shares held by funds that are distant from rating thresholds (the distance above the top 10^{th} (30^{th}) percentile value), all scaled by the number of shares outstanding. $Covariates_{i,t}$ are a vector of stock characteristics that include the logarithmic of market capitalization, logarithmic of book-to-market ratio, past 12-month cumulative stock return excluding the most recent month, past one-month return excluding the last trading day of the month, number of analysts covering the stock, and quarterly estimates of Gibbs estimate of effective costs proposed by Hasbrouck (2004) and squared-root version of Amihud (2002) measure of price impact. All independent variables are measured as of the second-to-last trading day of the quarter. In columns (3) and (6), the regression includes stock fixed-effects (α_i). All regressions include time fixed-effects (θ_t), standard errors are double-clustered by stock and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. The sample covers the period from 1990 to 2017.

Panel A: Using 10% cutoff to define funds that are "near" or "distant" from a rating threshold.

Dependent variable:	$R_t^{\text{Last Day}}$ (%)			$R_{t+1}^{\text{First Day}}$ (%)		
	(1)	(2)	(3)	(4)	(5)	(6)
Ownership$^{\text{Near-Distant}}$	1.82***	1.58***	1.35***	−1.49***	−1.19**	−0.91*
	(3.39)	(3.10)	(2.65)	(−2.94)	(−2.48)	(−1.79)
log(Market cap)		−0.09***	−0.09		0.05**	−0.27***
		(−3.33)	(−1.38)		(2.05)	(−3.69)
log(Book-to-market)		−0.04	−0.003		0.18***	0.10*
		(−1.33)	(−0.07)		(4.37)	(1.79)
One-month return		−1.40***	−1.39***		−0.27	−0.03
		(−4.91)	(−5.17)		(−0.62)	(−0.08)
Twelve-month return		0.13	0.08		−0.26	−0.17
		(1.56)	(0.97)		(−1.20)	(−0.85)
Analyst coverage		−0.001	−0.12**		−0.03	0.08
		(−0.05)	(−2.27)		(−0.86)	(1.51)
Gibbs' effective costs		0.40***	0.24*		−0.17	−0.08
		(2.95)	(1.88)		(−0.94)	(−0.59)
Amihud's price impact		−0.21	0.32		0.26	0.14
		(−0.74)	(1.07)		(0.66)	(0.38)
Time fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Stock fixed-effects	No	No	Yes	No	No	Yes
Observations	236,692	236,692	236,692	236,692	236,692	236,692
Adjusted R^2	0.14	0.14	0.17	0.21	0.21	0.23

Table 10–*Continued*

Panel B: Using 30% cutoff to define funds that are "near" or "distant" from a rating threshold.

Dependent variable:	$R_t^{\text{Last Day}}$ (%)			$R_{t+1}^{\text{First Day}}$ (%)		
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{Ownership}^{\text{Near}-\text{Distant}}$	0.52**	0.49*	0.39	−0.63**	−0.48	−0.32
	(2.01)	(1.93)	(1.61)	(−1.97)	(−1.55)	(−1.05)
log(Market cap)		−0.09***	−0.09		0.05**	−0.27***
		(−3.35)	(−1.40)		(2.06)	(−3.68)
log(Book-to-market)		−0.04	−0.004		0.18***	0.10*
		(−1.34)	(−0.07)		(4.37)	(1.79)
One-month return		−1.40***	−1.39***		−0.27	−0.03
		(−4.90)	(−5.17)		(−0.62)	(−0.08)
Twelve-month return		0.13	0.08		−0.26	−0.17
		(1.57)	(0.98)		(−1.20)	(−0.85)
Analyst coverage		−0.001	−0.12**		−0.02	0.08
		(−0.05)	(−2.27)		(−0.85)	(1.51)
Gibbs' effective costs		0.40***	0.24*		−0.17	−0.08
		(2.94)	(1.88)		(−0.94)	(−0.59)
Amihud's price impact		−0.21	0.32		0.26	0.14
		(−0.74)	(1.08)		(0.66)	(0.38)
Time fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Stock fixed-effects	No	No	Yes	No	No	Yes
Observations	236,692	236,692	236,692	236,692	236,692	236,692
Adjusted R^2	0.14	0.14	0.17	0.21	0.21	0.23

Table 11: Evidence from Stock Holdings: Placebo Tests

This table presents the results of the following linear regression model:

$$R_{i,t}^{Last\ day}\ (R_{i,t+1}^{First\ day}) = \beta \times Placebo\ ownership_{i,t}^{Near-Distant} + \gamma \times Covariates_{i,t}\ (+\alpha_t) + \theta_t + \varepsilon_{i,t}$$

where i indexes stocks and t indexes time in quarter. The dependent variable is stock i's return on the last trading day of quarter t in columns (1) through (3) and stock i's return on the first trading day of quarter $t+1$ in columns (4) through (6). The independent variable of interest, $Placebo\ ownership_{i,t}^{Near-Distant}$, is calculated as the difference between the number of shares held by funds that are *placebo* near a rating threshold (the *placebo* distance below the bottom 20$^{\text{th}}$ percentile value) and the number of shares held by funds that are *placebo* distant from rating thresholds (the *placebo* distance above the top 20$^{\text{th}}$ percentile value), all scaled by the number of shares outstanding. The *placebo* distance to a rating threshold is calculated by reversing the June 2002 change in the Morningstar rating methodology as in Table 4. $Covariates_{i,t}$ are a vector of stock characteristics that include the logarithmic of market capitalization, logarithmic of book-to-market ratio, past 12-month cumulative stock return excluding the most recent month, past one-month return excluding the last trading day of the month, number of analysts covering the stock, and quarterly estimates of Gibbs estimate of effective costs proposed by Hasbrouck (2004) and squared-root version of Amihud (2002) measure of price impact. All independent variables are measured as of the second-to-last trading day of the quarter. In columns (3) and (6), the regression includes stock fixed-effects (α_i). All regressions include time fixed-effects (θ_t), standard errors are double-clustered by stock and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. The sample covers the period from 1990 to 2017.

Dependent variable:	$R_t^{\text{Last Day}}$ (%)			$R_{t+1}^{\text{First Day}}$ (%)		
	(1)	(2)	(3)	(4)	(5)	(6)
Placebo ownership$^{\text{Near-Distant}}$	0.11	0.09	0.11	0.01	0.03	0.13
	(0.36)	(0.28)	(0.37)	(0.02)	(0.07)	(0.32)
log(Market cap)		−0.09***	−0.09		0.05**	−0.27***
		(−3.35)	(−1.40)		(2.07)	(−3.68)
log(Book-to-market)		−0.04	−0.004		0.18***	0.10*
		(−1.36)	(−0.08)		(4.39)	(1.80)
One-month return		−1.39***	−1.39***		−0.27	−0.03
		(−4.90)	(−5.16)		(−0.62)	(−0.08)
Twelve-month return		0.13	0.08		−0.26	−0.17
		(1.57)	(0.98)		(−1.20)	(−0.85)
Analyst coverage		−0.001	−0.12**		−0.03	0.08
		(−0.04)	(−2.27)		(−0.86)	(1.51)
Gibbs' effective costs		0.40***	0.24*		−0.17	−0.08
		(2.95)	(1.88)		(−0.94)	(−0.59)
Amihud's price impact		−0.21	0.32		0.26	0.14
		(−0.74)	(1.08)		(0.67)	(0.38)
Time fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Stock fixed-effects	No	No	Yes	No	No	Yes
Observations	236,692	236,692	236,692	236,692	236,692	236,692
Adjusted R^2	0.14	0.14	0.17	0.21	0.21	0.23

Table 12: Evidence from Stock Holdings: Heterogeneity Tests

This table presents the results of the following linear regression model:

$$R_{i,t}^{Last\ day} = \delta \times Ownership_{i,t}^{Near-Distant} \times Liquidity_{i,t} + \beta \times Ownership_{,i,t}^{Near-Distant} + \rho \times Liquidity_{i,t} + \gamma \times Covariates_{i,t}\ (+\alpha_t) + \theta_t + \varepsilon_{i,t}$$

where i indexes stocks and t indexes time in quarter. The dependent variable is stock i's return on the last trading day of quarter t. For measures of liquidity, $Liquidity_{i,t}$, I use the Gibbs estimate of effective costs proposed by Hasbrouck (2004) and squared-root version of Amihud (2002) measure of price impact. $Ownership_{i,t}^{Near-Distant}$ is calculated as the difference between the number of shares held by funds that are near a rating threshold (the distance below the bottom 20$^{\text{th}}$ percentile value) and the number of shares held by funds that are distant from rating thresholds (the distance above the top 20$^{\text{th}}$ percentile value), all scaled by the number of shares outstanding. $Covariates_{i,t}$ are a vector of stock characteristics that include the logarithmic of market capitalization, logarithmic of book-to-market ratio, past 12-month cumulative stock return excluding the most recent month, past one-month return excluding the last trading day of the month, number of analysts covering the stock, and quarterly estimates of Gibbs estimate of effective costs proposed by Hasbrouck (2004) and squared-root version of Amihud (2002) measure of price impact. All independent variables are measured as of the second-to-last trading day of the quarter. In columns (3) and (6), the regression includes stock fixed-effects (α_i). All regressions include time fixed-effects (θ_t), standard errors are double-clustered by stock and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. The sample covers the period from 1990 to 2017.

Dependent variable:	$R_t^{\text{Last Day}}$ (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
Ownership$^{\text{Near-Distant}}\times$ Effective costs	2.44**	2.60***	2.15**			
	(2.45)	(2.68)	(2.34)			
Ownership$^{\text{Near-Distant}}\times$ Price impact				7.69*	8.23**	5.88
				(1.85)	(2.01)	(1.52)
Ownership$^{\text{Near-Distant}}$	−0.39	−0.52	−0.43	0.44	0.31	0.32
	(−0.81)	(−1.10)	(−0.92)	(1.06)	(0.77)	(0.82)
Gibbs' effective costs	0.50***	0.39***	0.23*		0.40***	0.24*
	(4.04)	(2.88)	(1.83)		(2.95)	(1.88)
Amihud's price impact		−0.21	0.32	0.71***	−0.23	0.31
		(−0.72)	(1.08)	(2.87)	(−0.80)	(1.04)
log(Market cap)		−0.09***	−0.09		−0.09***	−0.09
		(−3.35)	(−1.40)		(−3.34)	(−1.39)
log(Book-to-market)		−0.04	−0.003		−0.04	−0.004
		(−1.32)	(−0.07)		(−1.35)	(−0.08)
One-month return		−1.40***	−1.40***		−1.40***	−1.39***
		(−4.92)	(−5.18)		(−4.91)	(−5.17)
Twelve-month return		0.13	0.07		0.13	0.08
		(1.56)	(0.97)		(1.56)	(0.97)
Analyst coverage		−0.002	−0.12**		−0.002	−0.12**
		(−0.07)	(−2.27)		(−0.09)	(−2.28)
Time fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Stock fixed-effects	No	No	Yes	No	No	Yes
Observations	236,692	236,692	236,692	236,692	236,692	236,692
Adjusted R^2	0.14	0.14	0.17	0.14	0.14	0.17

Table 13: Evidence from Stock Holdings: Intraday Returns

This table presents the results of the following linear regression model:

$$R_{i,t}^{Last\ intraday} = \beta \times Ownership_{i,t}^{Near-Distant} + \gamma \times Covariates_{i,t} + \theta_t + \varepsilon_{i,t}$$

where i indexes stocks and t indexes time in quarter. The dependent variable is stock i's half-hour intraday return on the last trading day of quarter t All intraday returns are calculated over an interval of thirty minutes from 9:00 to 16:00 using the NYSE's intrada Trade and Quote (TAQ) data. The independent variable of interest, $Ownership_{i,t}^{Near-Distant}$, is calculated as the difference between the number of shares held by funds that are near a rating threshold (the distance below the bottom 20$^{\text{th}}$ percentile value) and the number of shares held by funds that are distant from rating thresholds (the distance above the top 20$^{\text{th}}$ percentile value), all scaled by the number of shares outstanding. $Covariates_{i,t}$ are a vector of stock characteristics that include the logarithmic of market capitalization, logarithmic of book-to-market ratio, past 12-month cumulative stock return excluding the most recent month, past one-month return excluding the last trading day of the month, number of analysts covering the stock, and quarterly estimates of Gibbs estimate of effective costs proposed by Hasbrouck (2004) and squared-root version of Amihud (2002) measure of price impact. All independent variables are measured as of the second-to-last trading day of the quarter. All regressions include time fixed-effects (θ_t), standard errors are double-clustered by stock and time, and the resulting t-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. The sample covers year 2004 and the period from 2006 to 2014.

Dependent variable:	$R_{i,t}^{Last\ intraday}$ (%)												
Beginning time:	09:30	10:00	10:30	11:00	11:30	12:00	12:30	13:00	13:30	14:00	14:30	15:00	15:30
Ending time:	10:00	10:30	11:00	11:30	12:00	12:30	13:00	13:30	14:00	14:30	15:00	15:30	16:00
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
$\text{Ownership}^{\text{Near−Distant}}$	−0.06	0.15	−0.04	−0.13	−0.11	0.08	−0.07	−0.04	0.07	0.01	0.03	0.07	0.53***
	(−0.22)	(1.09)	(−0.43)	(−1.44)	(−1.45)	(1.10)	(−0.82)	(−0.63)	(0.94)	(0.12)	(0.39)	(0.76)	(3.79)
log(Market cap)	−0.01	0.01	−0.01	−0.02***	−0.01	0.01	−0.01	−0.01*	−0.003	0.002	0.005	0.01*	0.06***
	(−0.58)	(0.95)	(−0.81)	(−2.79)	(−1.05)	(0.87)	(−1.29)	(−1.83)	(−0.70)	(0.41)	(1.07)	(1.68)	(4.83)
log(Book-to-market)	−0.04**	0.001	−0.002	0.01	−0.001	0.01	0.002	−0.001	0.01*	0.01*	−0.01	−0.004	0.002
	(−1.99)	(0.09)	(−0.26)	(1.46)	(−0.32)	(0.87)	(0.55)	(−0.10)	(1.88)	(1.80)	(−1.06)	(−0.58)	(0.23)
One-month return	0.11	−0.20	−0.01	−0.19*	−0.10	−0.09	0.02	−0.05	−0.05	−0.06	−0.02	−0.10*	−0.08
	(0.35)	(−1.47)	(−0.22)	(−1.89)	(−0.89)	(−1.48)	(0.20)	(−0.74)	(−1.43)	(−1.04)	(−0.65)	(−1.81)	(−0.91)
Twelve-month return	0.09**	0.02	0.01	−0.01	0.01	−0.02*	−0.01	0.003	−0.02**	−0.01	−0.003	−0.01	−0.01
	(2.03)	(0.67)	(0.30)	(−0.83)	(0.34)	(−1.81)	(−1.61)	(0.23)	(−1.96)	(−0.96)	(−0.19)	(−0.41)	(−0.48)
Analyst Coverage	−0.002	−0.01	0.01	0.004	0.01	−0.004	−0.003	0.01	−0.003	0.002	0.01	0.002	−0.01
	(−0.11)	(−1.15)	(1.17)	(0.53)	(0.97)	(−0.87)	(−0.66)	(1.04)	(−0.79)	(0.55)	(1.58)	(0.26)	(−1.15)
Gibbs' effective costs	0.01	−0.03	0.02	0.03	0.01	0.02	0.03	0.04*	0.05	0.06*	0.05	0.004	0.003
	(0.11)	(−0.44)	(0.58)	(1.04)	(0.34)	(0.66)	(1.52)	(1.70)	(1.39)	(1.80)	(1.22)	(0.09)	(0.06)
Amihud's price impact	−0.30	−0.29	−0.08	−0.33**	−0.08	0.05	−0.29***	−0.04	−0.06	−0.04	0.31**	0.27	0.13
	(−0.84)	(−1.03)	(−0.42)	(−2.01)	(−0.92)	(0.39)	(−3.01)	(−0.20)	(−0.56)	(−0.33)	(2.17)	(1.30)	(0.55)
Time fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	85,600	85,600	85,600	85,600	85,600	85,600	85,600	85,600	85,600	85,600	85,600	85,600	85,600
Adjusted R^2	0.23	0.21	0.10	0.20	0.08	0.13	0.17	0.11	0.15	0.22	0.17	0.19	0.28