# Do Mutual Funds Manipulate Star Ratings? Evidence from Portfolio Pumping

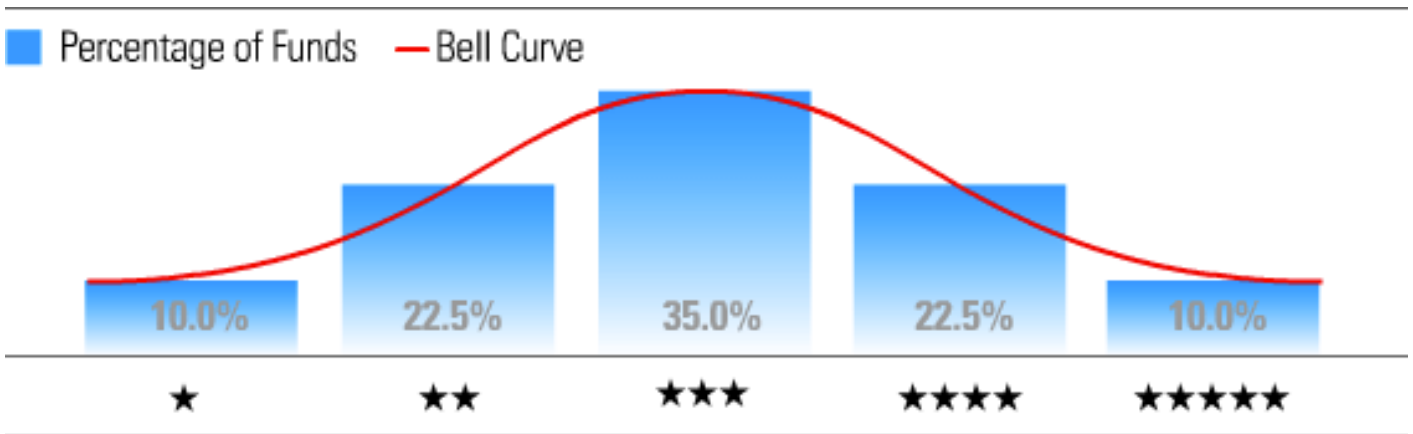
Sanghyun "Hugh" Kim

University of Texas at Dallas

December 31, 2020

# Morningstar Ratings

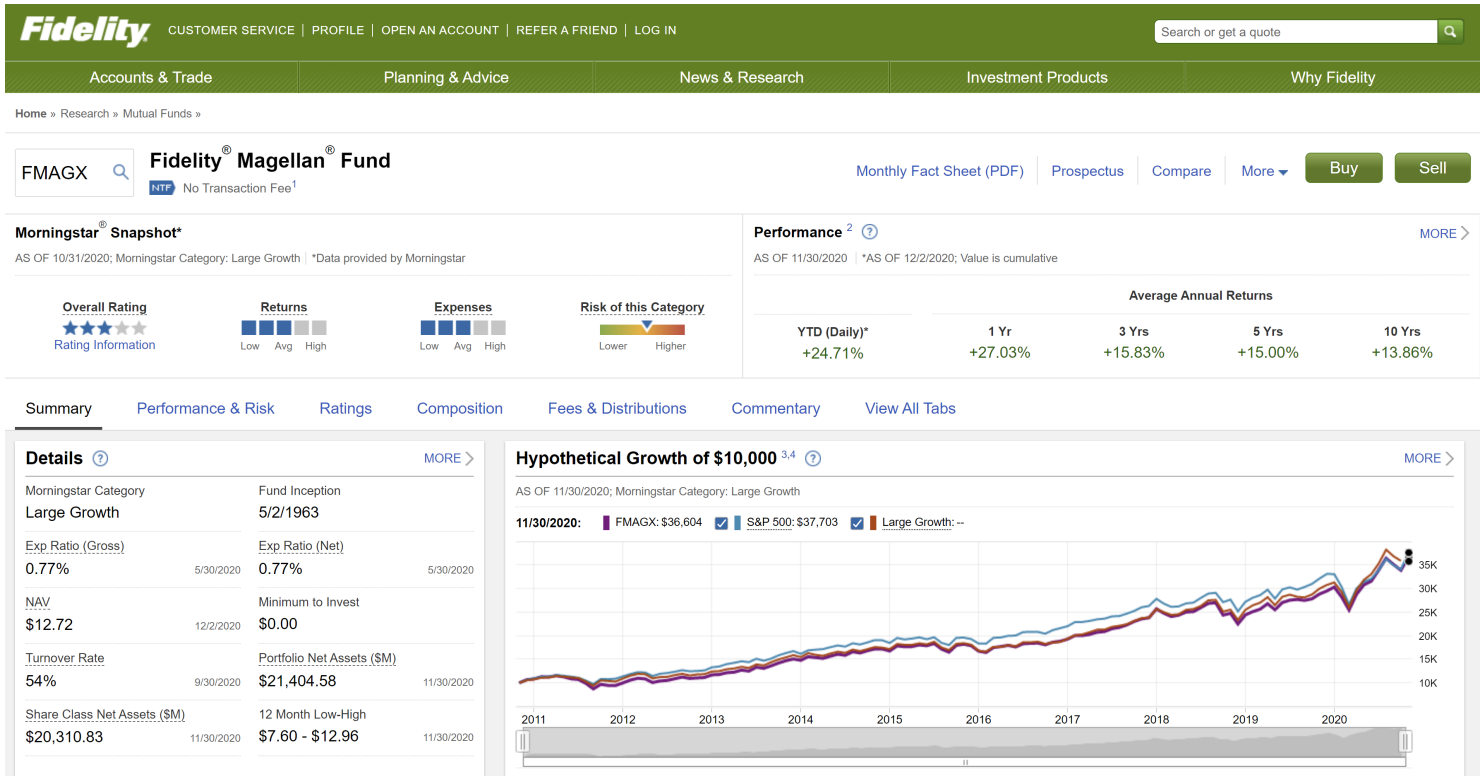
- Since its introduction in 1985, Morningstar's five-star rating system has become widely accepted in the mutual fund industry (Del Guercio and Tkac (2008)).
- At the end of every month, mutual fund share classes are rated on the basis of Morningstar Risk-Adjusted Returns (MRAR) on an integer scale of one star (lowest) to five stars (highest).




Source: Morningstar

# Morningstar Ratings

- Star ratings offer less sophisticated investors a simple and intuitive tool to use to allocate their capital across mutual funds.

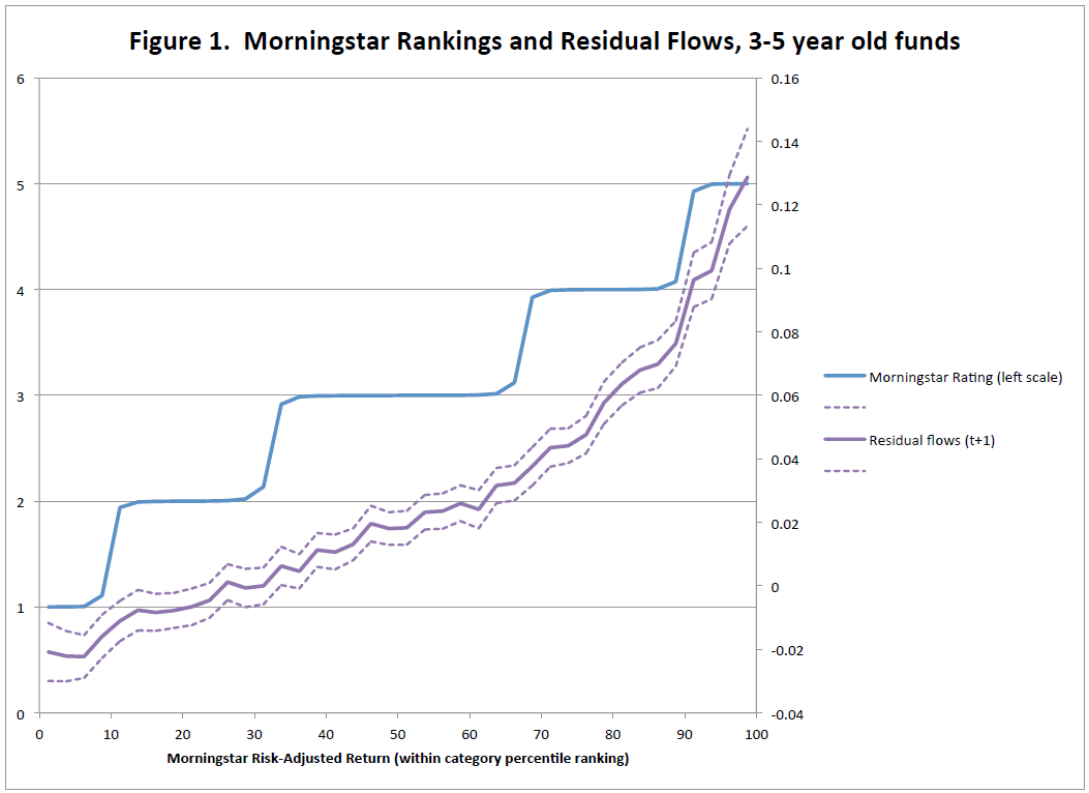


# Star Ratings on Fund Flows

- Discrete star ratings have a powerful influence on fund flows, independent of the underlying continuous performance measures (Del Guercio and Tkac (2008)).
- Mutual fund investors use simple heuristics such as star ratings, rather than asset-pricing models, for risk adjustment (Evans and Sun (Forthcoming)).
- Star ratings explain mutual fund investors' behavior much better than any asset pricing models (Ben-David et al. (2019)).

# Star Ratings on Fund Flows

- Reuter and Zitzewitz (2015) find that a large fraction of the difference in future fund flows received by five- and one-star funds represents a causal effect of the difference in star ratings on fund flows.




Source: Reuter and Zitzewitz (2015)

# My Paper

- Mutual funds manipulate star ratings by inflating their month-end portfolio values, especially when they are likely to finish the month near rating cutoffs.
  - Since open-end mutual funds calculate their net asset values (NAV) from the closing prices of their holdings, fund managers can artificially inflate the closing prices of their holdings by aggressively purchasing stocks they already own (Zweig (1997), Carhart et al. (2002)).

# Contribution to the Literature

- My paper is the first to link portfolio pumping to Morningstar's star ratings, contributing to several strands of literature.
  - Portfolio Pumping: Carhart et al. (2002), Ben-David et al. (2013), Hu et al. (2014), Duong and Meschke (2020), Patel and Sarkissian (Forthcoming)
  - Morningstar Ratings: Del Guercio and Tkac (2008), Reuter and Zitzewitz (2015), Evans and Sun (Forthcoming), Ben-David et al. (2019)
  - Managerial Incentives: Ippolito (1992), Sirri and Tufano (1998), Chevalier and Ellison (1997), Bollen and Pool (2009), Begley (2015), Lee et al. (2019)

# Star Ratings

- Star ratings are determined by within-category rankings of Morningstar Risk-Adjusted Returns (MRAR) over the prior three, five, and ten years, depending on data availability.

$$MRAR(\gamma) = \left[\frac{1}{T}\sum_{t=1}^{T}(1+ER_t)^{-\gamma}\right]^{-\frac{12}{\gamma}} - 1, \quad \gamma > -1,\ \gamma \neq 0$$

- $ER_t$ is the geometric excess return over the risk-free rate in month $t$.
- $\gamma$ is the risk-aversion coefficient.
- $T$ is the number of months in the time period.

# Distance to a Rating Threshold

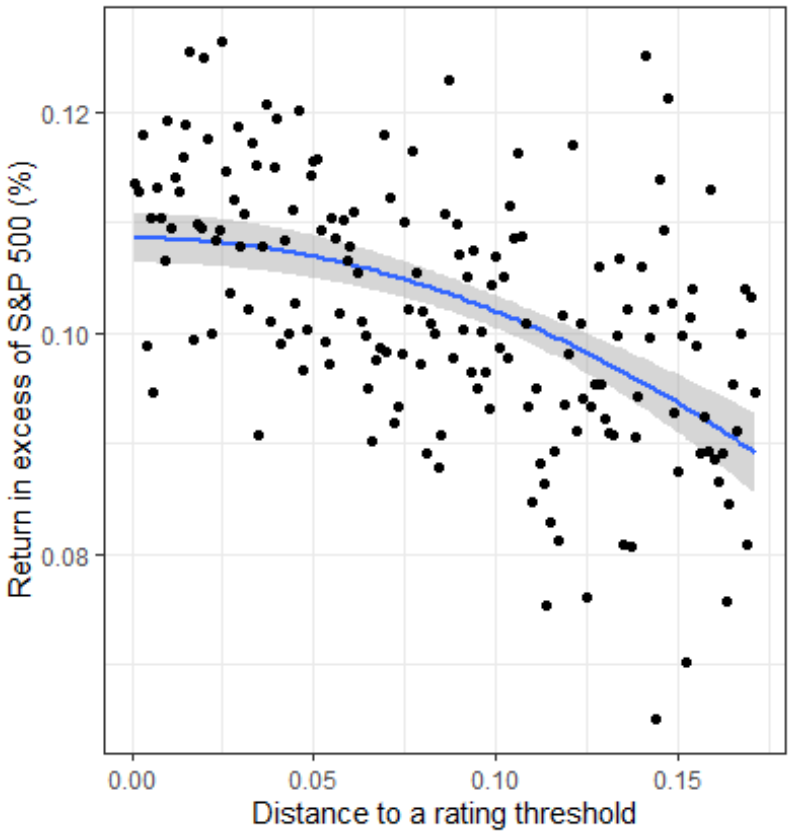
1. I estimate within-category MRAR rankings just prior to monthly rating updates, measured at the end of the *second-to-last* trading day of the month.

2. I compute the distance to a rating threshold as the distance between within-category percentile rankings and the nearest rating threshold.

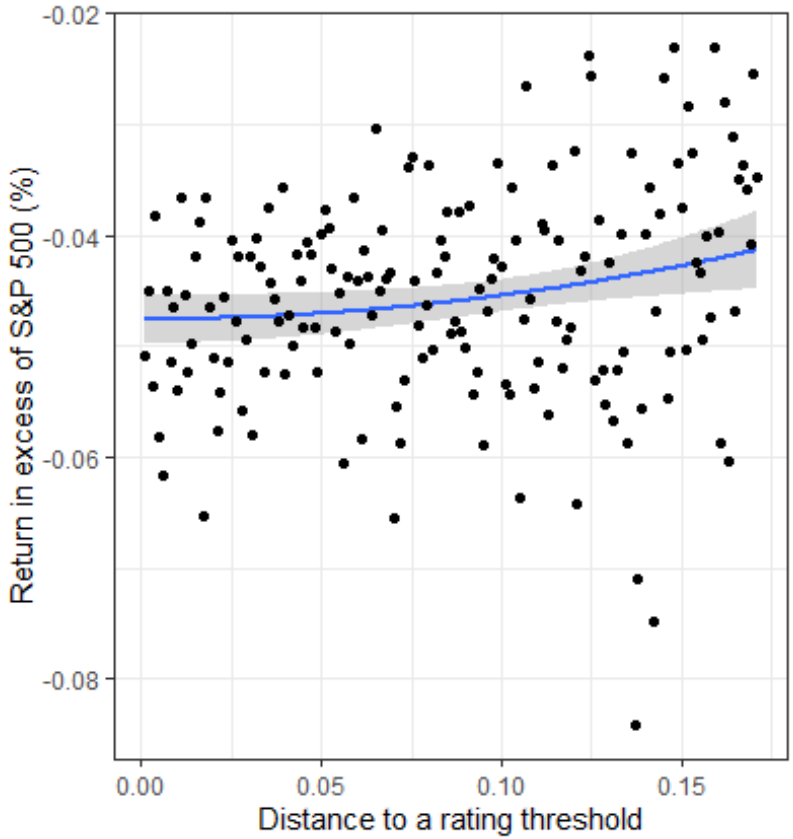
# Baseline Tests

- Whether funds near a rating threshold on the *second-to-last* trading day of the month would earn higher returns on the *last* trading day of the month, compared to funds that are farther away from rating thresholds?
  - Analogous to asking whether borderline A students near the end of the semester would perform better on the final exam, compared to students in the middle of A or B ranges?

## Distance to a Rating Threshold and Month-End Performance Inflation




Panel A: On the last day of month $t$




Panel B: On the first day of month $t+1$

## Distance to a Rating Threshold and Month-End Performance Inflation

$$R_{i,t}^{\text{Last day}} \ (R_{i,t+1}^{\text{First day}}) = \beta \times \text{Squared distance}_{i,t} + \gamma \times \text{Covariates}_{i,t} + \theta_{i,t} + \varepsilon_{i,t}$$

| | $R_t^{\text{Last Day}}$ | | $R_{t+1}^{\text{First day}}$ | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Squared distance | −0.44*** | −0.33*** | 0.24* | 0.19* |
| | (−4.00) | (−3.31) | (1.95) | (1.76) |
| Control variables | No | Yes | No | Yes |
| Category × Month FE | Yes | Yes | Yes | Yes |
| Observations | 1,252,358 | 1,110,321 | 1,252,358 | 1,110,321 |
| Adjusted $R^2$ | 0.45 | 0.46 | 0.45 | 0.46 |

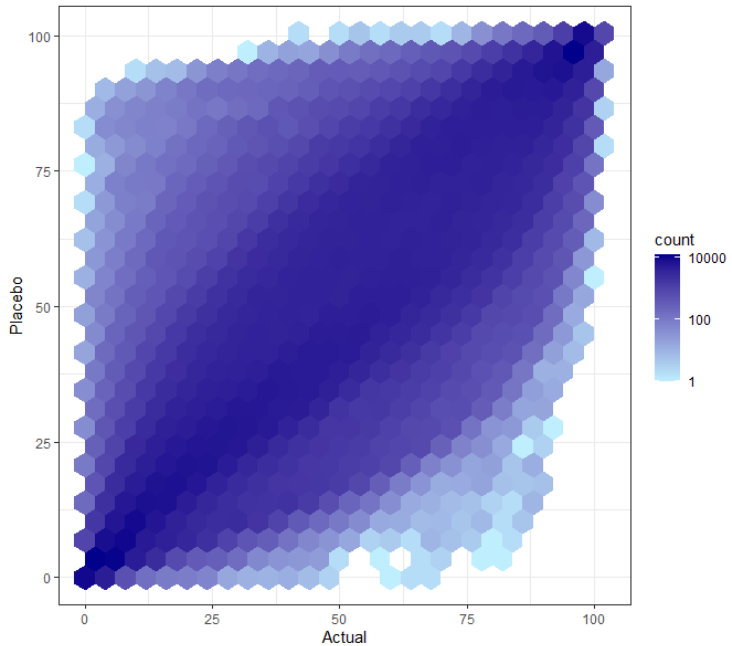
Standard errors are double-clustered by fund and by month, and the resulting t-statistics are reported in parentheses. The sample covers the period from 1990 to 2018.

# Placebo Tests

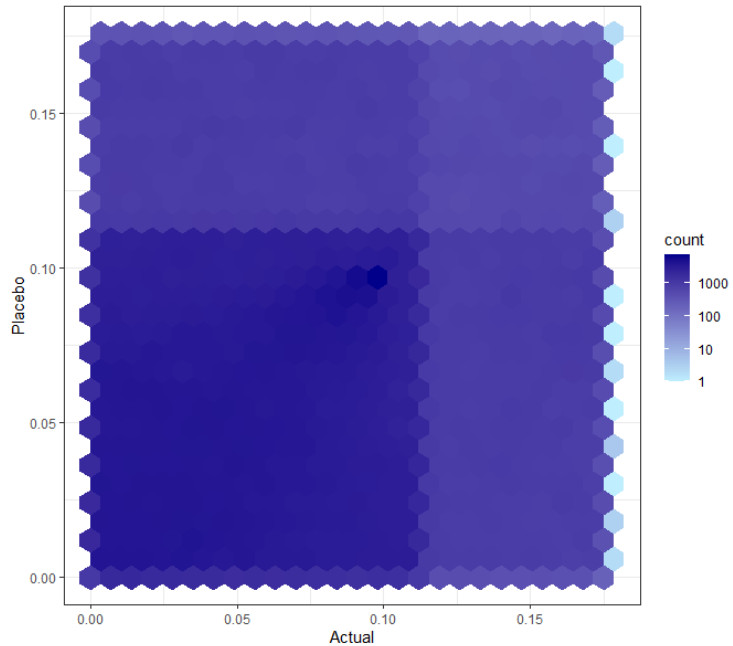
- To establish causality, I exploit major changes in Morningstar's rating methodology in June 2002.
- Along with the change in the risk adjustment process, Morningstar refined its peer groups used to rank mutual funds.
  - Morningstar started ranking U.S. equity mutual funds within its nine (three-by-three style box) categories along the size dimension (small, mid-cap, or large) and value dimension (value, blend, or growth).
  - All U.S. equity mutual funds, as a single category group, were ranked against each other prior to the change.
- I conduct placebo tests by reversing the June 2002 change.

# Placebo Tests

Correlation between Actual vs. Placebo Rankings and Distances to Rating Thresholds




Panel A: Within-category percentile rankings ($\rho = 0.76$)




Panel B: Distances to rating thresholds ($\rho = 0.10$)

## Placebo Tests: Reversing the June 2002 Change in Morningstar Rating Methodology

$$R_{i,t}^{\text{Last day}} \ (R_{i,t+1}^{\text{First day}}) = \beta \times \text{Squared placebo distance}_{i,t} + \gamma \times \text{Covariates}_{i,t-1} + \theta_{i,t} + \varepsilon_{i,t}$$

| | $R_t^{\text{Last Day}}$ | | $R_{t+1}^{\text{First day}}$ | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Squared *placebo* distance | 0.01 | 0.09 | 0.09 | 0.03 |
| | (0.13) | (1.03) | (1.10) | (0.39) |
| Control variables | No | Yes | No | Yes |
| Category × Month FE | Yes | Yes | Yes | Yes |
| Observations | 1,252,358 | 1,095,824 | 1,252,358 | 1,095,824 |
| Adjusted $R^2$ | 0.45 | 0.46 | 0.45 | 0.46 |

Standard errors are double-clustered by fund and by month, and the resulting t-statistics are reported in parentheses. The sample covers the period from 1990 to 2018.

# Cross-Sectional Tests I

- Not all star ratings are created equal!

1. Because the assignment of star ratings is subject to data availability, it becomes much more difficult for funds to manipulate star ratings as their return history extends further.

2. Discontinuities in the flow-performance relation are greater at higher rating cutoffs and strongest at the four/five-star cutoff (Reuter and Zitzewitz (2015)).

# Are All Star Ratings Created Equal?

$$R_{i,t}^{Last\ day} = \delta \times Squared\ distance_{i,t} \times Sensitivity_{i,t} + \beta \times Squared\ distance_{i,t} + \rho \times Sensitivity_{i,t} + \gamma \times Covariates_{i,t-1} + \theta_{i,t} + \varepsilon_{i,t}$$

| | $R_t^{\text{Last Day}}$ | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Squared distance $\times$ $\mathbb{1}$(Three-year rating) | $-0.52^{**}$ | $-0.55^{**}$ | | |
| | $(-2.21)$ | $(-2.36)$ | | |
| Squared distance $\times$ $\mathbb{1}$(Four/five-star cutoff) | | | $-1.38^{**}$ | $-1.10^{**}$ |
| | | | $(-2.48)$ | $(-1.98)$ |
| Squared distance | $-0.29^{***}$ | $-0.19^{**}$ | $-0.26^{**}$ | $-0.19$ |
| | $(-2.85)$ | $(-2.07)$ | $(-2.07)$ | $(-1.55)$ |
| $\mathbb{1}$(Three-year rating) | $0.01^{**}$ | $-0.004$ | | |
| | $(2.17)$ | $(-1.35)$ | | |
| $\mathbb{1}$(Four/five-star cutoff) | | | $0.02^{**}$ | $0.02^{*}$ |
| | | | $(2.08)$ | $(1.79)$ |
| Control variables | No | Yes | No | Yes |
| Category $\times$ Month FE | Yes | Yes | Yes | Yes |
| Observations | 1,252,358 | 1,095,824 | 1,252,358 | 1,095,824 |
| Adjusted $R^2$ | 0.45 | 0.46 | 0.45 | 0.46 |

Standard errors are double-clustered by fund and by month, and the resulting t-statistics are reported in parentheses. The sample covers the period from 1990 to 2018.

# Cross-Sectional Tests II

3. Portfolio pumping should be more pronounced among small-cap funds because the closing prices of less liquid stocks would presumably be easier to influence (Carhart et al. (2002)).

4. Peer effects among teams such as the presence of peer monitoring and joint monetary incentives are effective in deterring fund managers from engaging in illegal trading activities (Patel and Sarkissian (Forthcoming)).

# More Cross-Sectional Tests

$$R^{Last\ day}_{i,t} = \delta \times Squared\ distance_{i,t} \times Sensitivity_{i,t} + \beta \times Squared\ distance_{i,t} + \rho \times Sensitivity_{i,t} + \gamma \times Covariates_{i,t-1} + \theta_{i,t} + \varepsilon_{i,t}$$

| | $R^{\text{Last Day}}_t$ | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Squared distance × 𝟙(Small-cap) | −0.63** | −0.68** | | |
| | (−2.16) | (−2.35) | | |
| Squared distance × log(N Managers) | | | 0.16 | 0.16* |
| | | | (1.62) | (1.68) |
| Squared distance | −0.22* | −0.12 | −0.52*** | −0.44*** |
| | (−1.85) | (−1.11) | (−3.55) | (−3.31) |
| 𝟙(Small-cap) | 0.09*** | 0.08*** | | |
| | (3.82) | (3.61) | | |
| log(N Managers) | | | −0.0004 | −0.001 |
| | | | (−0.24) | (−0.30) |
| Control variables | No | Yes | No | Yes |
| Category × Month FE | No | No | Yes | Yes |
| Month FE | Yes | Yes | No | No |
| Observations | 1,252,358 | 1,095,824 | 1,239,838 | 1,085,804 |
| Adjusted $R^2$ | 0.29 | 0.30 | 0.46 | 0.47 |

Standard errors are double-clustered by fund and by month, and the resulting t-statistics are reported in parentheses. The sample covers the period from 1990 to 2018.

# Sub-Period Tests

- Following Carhart et al. (2002), the SEC started to investigate suspicious trading activities (Duong and Meschke (2020), Patel and Sarkissian (Forthcoming)).
- Portfolio pumping has become more evasive (Hu et al. (2014), Wang (2019)).

# Has Portfolio Pumping Become More Evasive?

$$R_{i,t}^{\text{Last day}} = \beta \times \textit{Squared distance}_{i,t} + \delta \times \textit{Squared distance}_{i,t} \times \textit{Quarter-end}_t + \gamma \times \textit{Covariates}_{i,t-1} + \theta_{i,t} + \varepsilon_{i,t}$$

| | $R_t^{\text{Last Day}}$ | | | |
|---|---|---|---|---|
| | 1990:01–2002:05 | | 2002:06–2018:12 | |
| | (1) | (2) | (3) | (4) |
| Squared distance × 𝟙(Quarter-end) | −1.34* | −1.51** | 0.22* | 0.24** |
| | (−1.79) | (−2.05) | (1.93) | (2.08) |
| Squared distance | −1.11** | −0.70* | −0.17** | −0.14* |
| | (−2.35) | (−1.70) | (−2.07) | (−1.86) |
| Control variables | No | Yes | No | Yes |
| Category × Month FE | Yes | Yes | Yes | Yes |
| Observations | 160,944 | 151,052 | 1,091,414 | 944,772 |
| Adjusted $R^2$ | 0.28 | 0.28 | 0.56 | 0.57 |

Standard errors are double-clustered by fund and by month, and the resulting t-statistics are reported in parentheses. The sample covers the period from 1990 to 2018 and is split around the June 2002 change in Morningstar's rating methodology.

# The Effect on Fund Flows

- I examine whether star rating manipulation through portfolio pumping is effective.
- I exploit a two-stage least squares (2SLS) estimation to show that month-end performance inflation
  1. increases the probability of a rating upgrade and
  2. increases fund flows in the month of a rating upgrade.

## The Effect of Portfolio Pumping on Star Ratings

$$\mathbb{1}(\textit{Ratings change}_{i,t+s}) = \beta \times \frac{R_t^{\text{Last Day}} - R_{t+1}^{\text{First Day}}}{2} + \gamma \times \textit{Covariates}_{i,t-1} + \theta_{i,t+s} + \varepsilon_{i,t+s}, \quad s = 0, 1$$

| *Dependent variable:* | $\mathbb{1}(\text{Upgrade}_t)$ | | $\mathbb{1}(\text{Downgrade}_{t+1} \mid \text{Upgrade}_t)$ | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $(R_t^{\text{Last Day}} - R_{t+1}^{\text{First Day}})/2$ | 0.02*** | 0.02*** | 0.09*** | 0.08*** |
| | (3.74) | (3.63) | (6.25) | (6.08) |
| Control variables | No | Yes | No | Yes |
| Category × Month FE | Yes | Yes | Yes | Yes |
| Observations | 1,241,336 | 1,087,734 | 86,391 | 75,597 |
| Adjusted $R^2$ | 0.01 | 0.01 | 0.06 | 0.06 |

Standard errors are double-clustered by fund and by month, and the resulting t-statistics are reported in parentheses. The sample covers the period from 1990 to 2018.

# The Effect of Star Rating Manipulation on Fund Flows

$$\mathbb{1}(Upgrade_{i,t}) = \beta_1 \times \frac{R_t^{\text{Last Day}} - R_{t+1}^{\text{First Day}}}{2} + \gamma_1 \times Covariates_{i,t-1} + \theta_{1,i,t} + \varepsilon_{1,i,t} \quad \text{(first stage)}$$

$$Flow_{i,t+s} = \beta_2 \times \widehat{\mathbb{1}(Upgrade_{i,t+1})} + \gamma_2 \times Covariates_{i,t-1} + \theta_{2,i,t+s} + \varepsilon_{2,i,t+s}, \quad s = 1, 2, 3 \quad \text{(second stage)}$$

| | First-stage | | Second-stage | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Dependent variable:* | $\mathbb{1}(\text{Upgrade}_t)$ | | $\text{Flow}_{t+1}$ | | $\text{Flow}_{t+2}$ | | $\text{Flow}_{t+3}$ | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $\widehat{\mathbb{1}(\text{Upgrade}_t)}$ | | | 8.54*** | 8.43*** | 2.87 | 2.97 | 3.27 | 3.52 |
| | | | (3.48) | (3.60) | (1.10) | (1.16) | (1.24) | (1.39) |
| $(R_t^{\text{Last Day}} - R_{t+1}^{\text{First Day}})/2$ | 0.03*** | 0.03*** | | | | | | |
| | (7.90) | (8.01) | | | | | | |
| $\text{Flow}_t$ | −0.0000 | 0.0002 | 0.16*** | −0.35*** | 0.14*** | −0.31*** | 0.12*** | −0.30*** |
| | (−0.31) | (0.94) | (26.47) | (−17.58) | (30.12) | (−18.51) | (29.94) | (−23.04) |
| $\text{Flow}_{t-1}$ | | −0.0003 | | 0.57*** | | 0.50*** | | 0.47*** |
| | | (−1.10) | | (24.64) | | (25.62) | | (29.96) |
| $\text{Flow}_{t-2}$ | | 0.0001 | | 0.13*** | | 0.11*** | | 0.09*** |
| | | (1.53) | | (31.50) | | (32.15) | | (30.17) |
| $R_t^{\text{ex Last Day}}$ | 0.03*** | 0.03*** | −0.07 | −0.09 | 0.10 | 0.08 | 0.09 | 0.06 |
| | (30.76) | (30.64) | (−0.90) | (−1.23) | (1.22) | (1.03) | (1.07) | (0.83) |
| $R_{t-1}$ | | 0.001 | | 0.12*** | | 0.14*** | | 0.12*** |
| | | (0.91) | | (8.69) | | (10.08) | | (8.97) |
| $R_{t-2}$ | | 0.001 | | 0.12*** | | 0.11*** | | 0.13*** |
| | | (1.11) | | (10.42) | | (8.54) | | (9.10) |
| Control variables | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Category × Month FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 1,075,491 | 1,071,857 | 1,069,355 | 1,066,029 | 1,063,102 | 1,059,987 | 1,057,854 | 1,054,814 |
| Adjusted $R^2$ | 0.05 | 0.05 | -0.03 | 0.01 | 0.04 | 0.07 | 0.03 | 0.05 |

Standard errors are double-clustered by fund and by month, and the resulting t-statistics are reported in parentheses. The sample covers the period from 1990 to 2018.

# Conclusion

- Morningstar ratings can distort the incentives of mutual fund managers.
- Specifically, mutual fund managers pump their portfolios to manipulate star ratings when they are likely to finish the month in the vicinity of rating cutoffs.
- Portfolio pumping can improve star ratings and increase fund flows, especially in the month of a rating upgrade.

*"'We are all in the gutter, but some of us are looking at the stars."*

*– Oscar Wilde*

# References I

Begley, Taylor A., 2015, The real costs of corporate credit ratings, *Working Paper* .

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* .

Bollen, Nicolas P.B., and Veronika K. Pool, 2009, Do hedge fund managers misreport returns? Evidence from the pooled distribution, *Journal of Finance* 64, 2257–2288.

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.

Chevalier, Judith, and Glenn Ellison, 1997, Risk taking by mutual funds as a response to incentives, *Journal of Political Economy* 105, 1167–1200.

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.

Duong, Truong X., and Felix Meschke, 2020, The rise and fall of portfolio pumping among U.S. mutual funds, *Journal of Corporate Finance* 60.

Evans, Richard B, and Yang Sun, Forthcoming, Models or stars: The role of asset pricing models and heuristics in investor risk adjustment, *Review of Financial Studies* .

# References II

Hu, Gang, R David Mclean, Jeffrey Pontiff, and Qinghai Wang, 2014, The year-end trading activities of institutional investors: Evidence from daily trades, *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.

Lee, Jung Hoon, Charles Trzcinka, and Shyam Venkatesan, 2019, Do portfolio manager contracts contract portfolio management?, *Journal of Finance* 74, 2543–2577.

Patel, Saurin, and Sergei Sarkissian, Forthcoming, Portfolio pumping and managerial structure, *Review of Financial Studies* .

Reuter, Jonathan, and Eric Zitzewitz, 2015, How much does size erode mutual fund performance? A regression discontinuity approach, *Working Paper* .

Sirri, Erik R, and Peter Tufano, 1998, Costly search and mutual fund flows, *Journal of Finance* 53, 1589–1622.

Wang, Pingle, 2019, Portfolio pumping in mutual fund families, *Working Paper* .

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