Is Mutual Fund Family Retirement Money Smart?

Pramodkumar Yadav[1]

Lebow College of Business, Drexel University

Abstract

Using data on investments of fund family employees in their 401(k) plans, I show that employee flows predict fund performance up to two years. The predictive power is stronger when fund family employees are located close to fund managers, pointing to employees exploiting their proximity to managers to learn about the managers' skill. The results are not driven by plan design, portfolio managers' ownership, or cross-subsidization. The top quintile of funds in terms of employee flows outperforms the bottom quintile by 1.6% annually in terms of Carhart Alpha, suggesting that other investors can benefit by mimicking fund employees.

JEL Classifications: G11, G14, G23

Keywords: Mutual Funds, Smart Money, Managerial Skill, Defined Contribution Pension Plans, Retirement Savings

[1] Email: py67@drexel.edu I am deeply indebted to Daniel Dorn for his invaluable guidance and discussions. I would like to thank Naveen Daniel, Caitlin Dannhauser, Umit Gurun, Markus Ibert, Zoran Ivkovich, Michelle Lowry, Jose Vincente Martinez, Mike Pagano, Haibei Zhao, and seminar participants at Drexel University for helpful comments.

The question of whether money is smart, that is, whether investors trade into funds that subsequently outperform, reflects the joint hypothesis of (i) some fund managers having skill and (ii) investors being able to detect that skill. Existing studies that examine the smart money question focus on fund flows aggregated across U.S. investors (Gruber 1996; Zheng 1999; Sapp and Tiwari, 2004; Frazzini and Lamont, 2008), U.K. individuals and institutions (Keswani and Stolin, 2008), and U.S. retirement plan participants (Sialm, Starks, and Zhang, 2015).

In this paper, I ask a targeted smart money question: do fund flows from fund family employees forecast future fund performance? My premise is that fund family employees are well placed to observe signals about managerial skill unavailable to the investing public. For employees such as fund family executives, fellow managers, analysts, or traders, these signals may come from direct observation of the talent or work ethic of managers. Other fund employees may not observe these signals directly, but can learn from their better placed co-workers (see Duflo and Saez, 2003 and Ouimet and Tate, 2019 for examples of other financial decisions that are influenced by co-workers).

I document that flows from fund employees indeed predict risk-adjusted fund performance and do so for up to two years – in contrast, prior papers find that predictive power of aggregate flows is limited to short horizons and reverts at two years (Lou, 2012). This finding prompts two additional research questions. First, what are employees smart about? For example, their investment actions may reflect information about managerial skill or advance knowledge about fund families' actions such as cross-subsidization. Second, can outside investors benefit from mimicking fund family employee investments? My contribution is to show that investments of fund family employees contain information about managerial skill, and this information can be exploited by other investors.

My analysis is based on a hand-collected data set of 401(k) holdings of mutual fund family employees. From 2009 onwards, these holdings can be extracted annually from the Form 5500 "Annual Return/Report of Employee Benefit Plan" that plan sponsors with more than 100 employees have to file with the Department of Labor. My data set spans actively managed domestic U.S. equity funds in the 401(k) plans of a sample of 34 fund families that account for approximately 80% of total net assets in the U.S. equity universe of the CRSP mutual fund database from 2009 to 2016.

To measure the propensity of fund family employees to actively trade into and out of funds in their retirement plan, I define *Employee Flows* in a given fund and year as the actual change in employees' investment in the fund (employee assets) during the year, minus the appreciation in the employee assets due to the performance of the fund's holdings during the year, scaled by the employee assets in the fund at the beginning of the year. The definition of *Employee Flows* is thus analogous to that of aggregate flows employed in the literature (see, e.g., Sirri and Tufano, 1998), but using employee assets instead of the funds' total net assets. In principle, flows to funds in 401(k) plans of fund families are jointly determined by fund families (the plan sponsors who select the funds in the plan) and investment allocations by the fund family employees (the plan participants who invest across the available options). Sialm, Starks, and Zhang (2015) report that plan sponsors actively adjust retirement plan menus by adding and dropping funds. I limit the sample for a given plan and year to funds that were available in that plan prior to that year and are still available in that plan at the end of that year. This ensures that *Employee Flows* reflect actions of participants rather than sponsors.[2]

[2] The definition of *Employee Flows* thus resembles that of "participant flows" in Sialm, Starks, and Zhang (2015). Sialm, Starks, and Zhang (2015), however, aggregate participant flows in a given fund across all retirement plans that include that fund, whereas I only consider the retirement plan of the fund family sponsor.

Employee Flows predict future gross and net alphas over horizons of up to two years. All else equal, a one-standard-deviation increase in *Employee Flows* in a given fund and year is associated with a significant increase of about 21 bp in Carhart (four-factor) net alpha during the following year and a cumulative increase of 36 bp during the following two years. The predictive power of *Employee Flows* is robust to controlling for a wide range of skill measures and fund attributes that have been linked to future alpha such as aggregate flows, expense ratio, past alpha (Bollen and Busse, 2005), Active Share (Cremers and Petajisto, 2009), R^2 (Amihud and Goyenko, 2013), Return Gap (Kacperczyk, Sialm, and Zheng, 2008), Tracking Error (Cremers and Petajisto, 2009), and Abnormal Turnover (Pastor, Stambaugh, and Taylor, 2017).

The long-term predictive power of *Employee Flows* is remarkable given that existing skill measures and aggregate flows have been reported to predict alphas mostly at horizons of one to three months. Wermers (2003) and Lou (2012) conjecture that the short-term predictive power of flows, in particular, may be driven by flow-induced price pressure rather than information about managerial skill. Consistent with this conjecture, Frazzini and Lamont (2008) and Lou (2012) report that the predictive power of aggregate flows reverses at longer horizons. Price pressure is an unlikely explanation for my results, both because *Employee Flows* predict longer-term performance and because *Employee Flows* are typically much smaller in dollar terms than aggregate flows.

The predictive power of *Employee Flows* is robust to alternative ways of defining *Employee Flows* (e.g., in dollar terms instead of scaling by employee assets) and future fund performance (e.g., CAPM alpha or raw return instead of Carhart 4-factor alpha). The results are also robust to controlling for highly granular family × year × style fixed effects.

Having established the predictive power of *Employee Flows*, I examine different explanations for the observed predictability. If proximity to fund managers facilitated the transmission of information regarding manager skill across plan participants, one would expect the predictive power of *Employee Flows* to be greater when managers are closer to plan participants. I exploit variation in the organizational and geographic distance between fund managers and plan participants to test this information transmission explanation. Typically, 401(k) menus of fund families consist of a mix of funds advised by the fund family that sponsors the plan ("internal funds") and funds advised or sub-advised by other fund families ("external funds"). Presumably, fund family employees can gather information more easily about fund managers in their own organization than about fund managers in other organizations. I classify a fund as internal versus external based on the sub-advisory structure reported in the fund's NSAR filings. Consistent with the information transmission explanation, variation in *Employee Flows* significantly predicts variation in future performance across internal funds, but not across external funds.

The geographic organization of fund families offers a second distance experiment. Portfolio managers of fund families operate either in a single location or in multiple locations. Presumably, operating in a single location facilitates information transmission within the organization. Thus, one would expect the predictive power of *Employee Flows* for the performance of internal funds to be stronger for single-location families. I classify fund families in my sample as single- vs multi-location based on the address information in their funds' NSAR filings. Consistent with the information transmission explanation, variation in *Employee Flows* significantly predicts variation in future performance across internal funds for single-location families, but not for internal funds for multi-location families.

I consider three alternative explanations for the observed predictability: *Employee Flows* could reflect (i) plan design, (ii) portfolio managers' ownership, or (iii) advance knowledge of the family's cross-subsidization efforts.

One alternative explanation is that fund families tilt their plans towards well-performing funds or somehow direct their employees towards such funds. I find that In-Plan-Funds (funds that appear in a fund family's 401(k) plan in the sample) do not outperform their Not-In-Plan peers (funds advised by a fund family in the sample, but not included in any of the sample's 401(k) plans). Moreover, 19 out of the sample's 34 fund families have been sued by their employees regarding their 401(k) plan offerings during the sample period. Finally, when fund families use default options, they tend to direct employee investments towards target-date or index funds as opposed to actively managed funds. All three observations are at odds with fund family actions driving the observed predictability.

Another explanation for the observed predictability is that *Employee Flows* reflect the portfolio managers' investments in their own funds. Khorana, Servaes, and Wedge (2007) and Ibert (2019) report that higher ownership by portfolio managers is associated with better future performance of the corresponding funds. I manually collect information about managerial ownership from the funds' Statements of Additional Information (SAIs). I do find that higher managerial ownership is associated with higher future fund alpha in some specifications. The correlation between managerial ownership and Employee Flows is virtually zero[3], however, and controlling for managerial ownership does not affect the predictive power of *Employee Flows*. Moreover,

[3] The low correlation is likely due to a combination of two factors: (i) only 40% of fund families report that fund managers hold stakes in their fund through their retirement, and (ii) 20% of the funds in the sample do not report any ownership by the portfolio manager.

managerial ownership cannot explain why *Employee Flows* predict the performance of internal funds in single-location families, but not in multi-location families.

The third alternative explanation I consider is that *Employee Flows* capture fund family employees' advance knowledge of a family's cross-subsidization strategy. Gaspar, Massa and Matos (2006) report that fund families might use allocations in underpriced IPOs to favor certain funds. *Employee Flows*, however, also predict future fund alpha when I correct alpha by estimating the return contribution of IPO allocations following the method of Gaspar, Massa and Matos (2006). Chaudhuri, Ivković, and Trzcinka (2018) argue that a family's cross-subsidization efforts will likely focus on funds that are both strong recent performers and relatively small in their family. Controlling for the interaction between past performance and relative fund size does not affect the predictive power of *Employee Flows*, either.

Finally, I address the question of whether other investors can benefit by mimicking *Employee Flows*, that is, trading into funds with large positive *Employee Flows* or trading out of (or avoiding) funds with large negative *Employee Flows*. Although fund families have to provide information about the holdings of their own 401(k) plan participants as of the end of each plan year, they only have to file this information with the Department of Labor (DOL) by the end of October of the following year. I consider the following feasible trading strategy. At the end of each plan-year, identify the funds that are in the top quintile in terms of *Employee Flows* during that plan-year. Next, buy an equally-weighted portfolio of such top-quintile funds at the end of October of the following year and holds this portfolio for 12 months. The top quintile of funds in terms of *Employee Flows* significantly outperforms the bottom quintile by 1.6 percent per year in terms of Carhart Alpha.[4] Moreover, the top quintile of funds in terms of *Employee Flows* significantly

[4] The performance differential of 1.6% between the top and bottom fund quintile in terms of *Employee Flows* appears to be substantially greater than the estimated increase in future Carhart Alpha of 21bp associated with a one-standard

outperforms an equally-weighted portfolio of all actively managed U.S. equity funds offered by the sample families (whether or not these funds are in a family's retirement plan) by 0.9 percent per year in terms of Carhart Alpha.

This paper contributes to the smart money literature and the literature on mutual fund skill. My study is the first to document that flows of fund family employees are smart. The analysis, which combines hand-collected data on employee investments and detailed fund information, allows me to distinguish between different mechanisms behind the observed smartness and identify information about managerial skill as the likely mechanism. The most closely related smart money paper is Sialm, Starks, and Zhang (2015) who aggregate flows across all DC plan participants. Unlike *Employee Flows*, aggregate DC flows are not smart. This is also consistent with the proximity-based mechanism proposed in this paper, as the distance between plan participants and the corresponding fund managers is generally large (though not just in geographic terms).

My study also adds to our understanding of where information about the skill of a mutual fund manager resides. Berk, van Binsbergen, and Liu (2017) report evidence that fund family executives know about the skill of their family's fund managers and use that knowledge to make value-enhancing decisions for the family. Specifically, they report that fund managers who are internally promoted by their family subsequently generate higher value for the family in terms of *dollar* alpha. In contrast, funds that receive higher *Employee Flows* generate higher *percentage* alpha. This difference reflects the differential incentives of fund family employees who care about

---

deviation increase in *Employee Flows* reported earlier. This difference can be explained by (i) the distance between the top and bottom quintiles of funds in terms of *Employee Flows* being larger than one standard deviation and (ii) the trading strategy being based on a univariate rather than a multi-variate analysis. Moreover, the trading strategy accounts for the delay in information disclosure. Consistent with the long-term predictive power of *Employee Flows*, the delay does not adversely effect the results.

percentage alpha as opposed to dollar alpha when it comes to their retirement investments. Viewed through the lens of Berk and Green (2004), the results suggest that fund family employees have an information advantage when judging the fund's actual size relative to its optimal size (which is determined by the manager's skill).

The overall take-away from the paper is that *Employee Flows* are informative about managerial skill and that investors can use this information to identify actively managed funds that perform relatively well in the future.

I. Data

This section describes the construction of the data set, defines key variables, and reports summary statistics for all variables.

A. Data Sources

The paper's data set results from merging information from five data sources. First, fund holdings in fund family 401(k) plans, extracted from Form 5500 "Annual Return/Report of Employee Benefit Plan," form the core of the paper's sample. Second, the Center for Research in Security Prices's (CRSP) Survivor-Bias-Free U.S. Mutual Fund Database (MFDB) supplies returns as well as other attributes for these funds. Third, NSAR filings provide detailed information on the identity of the funds' advisors, subadvisors, and their geographic location. Fourth, the funds' Statements of Additional Information (SAI) contain information about fund managers' holdings in these funds. Fifth, the Securities Data Corporation's Platinum database supplies information on recent initial public offerings (IPOs) held by these funds.

A.1 Form 5500 "Annual Return/Report of Employee Benefit Plan"

The data set is based on manually collecting 401(k) holdings of mutual fund families from the Form 5500 "Annual Return/Report of Employee Benefit Plan" from 2009 to 2016. Firms with more than 100 active plan participants at the beginning of each plan year are required to file this form with the Department of Labor (DOL). The form reports details of a firm's retirement plan as of the end of each plan year and is publicly available via the DOL's website (https://www.dol.gov/agencies/ebsa/about-ebsa/our-activities/public-disclosure/foia/form-5500-datasets). From 2009 onwards, form submissions include the so-called 'Schedule of Assets' which, at a minimum, lists the names of the plan's investment options and the dollar amount invested in each option as of the end of a calendar year. Smaller fund families with fewer than 100 plan participants tend to file Form 5500-SF (short form) which often does not include a Schedule of Assets. The availability of the Schedule of Assets determines the sample period and guides sample selection.

Mutual fund families are not required to report the asset holdings in their Schedule of Assets with standardized investment identifiers (such as CUSIP) or in a machine-readable format. To balance the effort of data collection with the goal of capturing as much of the fund universe as possible, I focus on the largest fund families. Specifically, in each year of the sample period, I identify the largest fund families whose funds collectively account for 80% of the total net assets (TNA) of domestic equity funds in the Center for Research in Security Prices's (CRSP) Survivor-Bias-Free U.S. Mutual Fund Database (MFDB). If a fund family meets the 80% cutoff in a year of the sample period, it is included for the entire sample period.

The Schedule of Assets for a given year includes information on all investment options with positive plan balances (e.g., registered investment advisory/mutual funds, collective investment trusts, separate accounts) and the dollar amount invested in each investment option as of the end of that year. Mutual funds account for 83% of assets across plan-years, on average.

A.2 CRSP MFDB

I assign a CRSP MFDB identifier to each fund on the Schedule of Assets on the basis of a Levenshtein edit distance based fuzzy match (Levenshtein 1965) between the name of the fund given on the Schedule of Assets and the fund name in the CRSP MFDB; this matching algorithm has been widely used (see, e.g., Mann, 2018). I verify all matches manually. I can identify 98% of the mutual funds in the plans corresponding to 99% of the plan balance in mutual funds.

Like Sialm, Starks, and Zhang (2015), I focus on domestic U.S. equity funds, the category with the most funds in the retirement plan of fund families. Domestic equity funds are identified on the basis of their CRSP objective codes beginning with EDC or EDY. The sample excludes hedged funds (beginning with code EDYH) as well as funds with a short bias (beginning with code EDYS). Index funds are identified manually (if the name includes index/ idx/ indx/ ETF) as are target date funds (if the name includes a target year). Domestic equity funds account for 60% of plan assets in mutual funds in the paper's sample, on average.

The Schedule of Assets does not consistently identify share classes, thus the match is at the fund level rather than at the share class level. I transform share class-level attributes into fund-level attributes by value-weighting the share class-level attributes using share class TNA as weights. The oldest share class determines fund age. I follow Evans (2010) and exclude funds with fewer than 36 monthly return observations and funds with TNA less than 5 million dollars to mitigate the effect of incubation bias. The resulting sample contains 1761 plan-fund-year observations representing 34 fund family plans and 336 distinct actively managed funds.

Fund families typically do not include all of the family's domestic equity funds as part of their 401(k) plan. For a given family and year, I identify "not-in-plan" funds as domestic equity funds that are advised by that family as per the CRSP MFDB, but are not part of the retirement plan menu of that family for the year. Not-in-plan funds account for 70% of the number of family funds and 61% of family fund TNA across families and over time, on average.

A.3 NSAR Filings

More than two thirds of fund families in the sample offer funds directly advised by the sponsoring family (to which I refer as "internal" funds) as well as funds that are advised or sub-advised by other fund families ("external" funds). A fund is classified as internal for a given fund family only if the sponsoring fund family serves as the sole advisor to the fund. To obtain information on fund sub-advisory details and the geographic location of fund managers, I rely on the semi-annual NSAR filings required by the Securities and Exchange Commission (SEC). Historical NSAR filings are downloaded from Edgar and merged with the existing data set using the fuzzy matching technique described above. The matching technique is able to identify 95% of the mutual funds in the sample (corresponding to 98% of the underlying plan balance).

A.4 Statements of Additional Information (SAI)

I manually extract information on managerial ownership from the funds' annual SAIs. Mutual funds are required to report managerial ownership using the following categories: $0, $1-$10,000, $10,001-$50,000, $50,001-$100,000, $100,001-$500,000, $500,001-$1,000,000, or above $1,000,000. Following Khorana, Servaes, and Wedge (2007), I use the mid-point of a reported category to estimate the manager's dollar ownership in his fund (using \$0 for the lowest category and \$1,000,000 for the highest category).

A.5 SDC Platinum

To estimate the contribution of IPO allocations to fund returns, I follow the methodology of Gaspar, Massa and Matos (2006). I extract all IPO deals from the Securities Data Company's (SDC) database from 2009 to 2017 and merge this information with stock prices from CRSP (to compute first-day underpricing) and with fund level holdings data in CRSP's MFDB.

B. Variable Construction

This Section discusses the definition and construction of the key independent variable (*Employee Flows*) and the key dependent variable (*Carhart Alpha*). The Appendix offers a detailed description of all the other variables used in the paper.

B.1 Employee Flows

To measure the propensity of fund family employees to actively trade into and out of funds in their retirement plan, I define *Employee Flows* as:

$$Employee\ Flows_{i,j,t} \equiv \frac{Employee\ Assets_{i,j,t} - Employee\ Assets_{i,j,t-1}(1+r_{i,t})}{Employee\ Assets_{i,j,t-1}} \qquad (1)$$

$subject\ to\ Employee\ Assets_{i,j,t-1} > 0$ and $Employee\ Assets_{i,j,t} > 0$

where *i* indexes fund, *j* indexes sponsor fund family, and *t* indexes year. $Employee\ Assets_{i,j,t}$ is the dollar market value of plan balance of fund *i* in the plan of fund family *j* at the end of year *t*. $r_{i,t}$ is return of fund *i* in year *t*. The definition of *Employee Flows* is thus analogous to that of aggregate flows employed in the literature (see, e.g., Sirri and Tufano 1998), but using employee assets instead of the funds' total net assets (TNA). I require $Employee\ Assets_{i,j,t-1} > 0$ and $Employee\ Assets_{i,j,t} > 0$, that is, a fund has to be available in the plan prior to year *t* and still be available in the plan at the end of year *t*. This requirement ensures that *Employee Flows* are driven by employees as opposed to the sponsoring fund family driving flows by adding funds to or deleting funds from the plan menu. The resulting *Employee Flows* variable is thus similar to "participant flows" as defined in Sialm, Starks, and Zhang (2015), but at the level of a specific retirement plan (that of the sponsoring fund family) rather than aggregated across a large sample of retirement plans.

I scale dollar employee flows by employee assets to get a percentage flow measure. The percentage measure reflects the fact that employees face a constrained optimization problem and are concerned about both fund alpha and idiosyncratic risk when choosing funds from their plan. For example, employees may expect a high alpha from a fund specializing in deep value stocks, but because deep value stocks represent a small fraction of their investment universe, they will only allocate a relatively small dollar amount to that fund. I use employee assets to scale flows rather than total net assets (TNA) because employee assets reflect the investment constraints faced by the employees in the plan whereas TNA reflects unrelated factors such as fund age and the fund family's distribution strategy. In robustness checks, I consider different definitions of *Employee Flows*, including unscaled dollar flows.

By defining non-employee assets for a given fund and time as the difference between a fund's TNA and its employee assets at that time, the corresponding definition of *Non-Employee Flows* emerges naturally as

$$Non\text{-}Employee\ Flows_{i,j,t} \equiv \frac{Non\text{-}Employee\ Assets_{i,j,t} - Non\text{-}Employee\ Assets_{i,j,t-1}(1+r_{i,t})}{Non\text{-}Employee\ Assets_{i,j,t-1}} \qquad (2)$$

To mitigate the impact of outliers, I winsorize all fund flows measures at the 1% and 99% levels.

B.2 Fund Performance

I use the Carhart's (1997) four-factor alpha as the baseline measure for fund performance. *Carhart Alpha* is estimated using a rolling-window approach (as in, e.g., Amihud and Goyenko, 2012). Specifically, factor betas in month *t* are estimated by regressing monthly fund returns in excess of the one-month T-bill rate on monthly factor realizations during the period *t-1* and *t-36*. Funds need to have a minimum of 24 months of data for the factor beta estimates (and consequently, for the alpha estimates) to be valid. The estimated factor betas are multiplied by the corresponding factor realizations in month *t* to compute a predicted excess return for month *t*. Predicted excess returns and realized excess returns are compounded over a year (or longer periods). *Carhart Alpha* is the difference between the compounded realized excess return and the compounded predicted excess return.

C. Summary Statistics

Table 1 reports summary statistics for measures of fund alpha and fund flows across the 1,761 fund-years in the sample, which consists of 34 fund families and a total of 336 distinct funds from 2009 to 2016. *Employee Flows* average 17%, indicating that actively managed funds in fund family retirement plans experienced substantial net inflows during the sample period. Non-employee flows and aggregate flows have lower means than *Employee Flows*, reflecting the fact that non-employee assets are typically much larger than employee assets; employee assets average 2% of TNA. *Employee Flows* are more volatile than non-employee flows and aggregate flows, consistent with fund employees being particularly active investors. *Carhart Alpha* is substantially negative during the sample period, averaging -144bp. Net returns average 12.4%, a reflection of the sustained bull market during the sample period.

II. *Employee Flows* Predict Future Fund Performance

This section documents the predictive power of *Employee Flows* for fund performance and shows that this predictive power is long-lived as well as robust to alternative definitions of *Employee Flows* and fund performance.

A. Baseline Results

To examine the predictive power of *Employee Flows,* I estimate the following panel regression model:

$$Performance_{i,j,t+1} = \gamma \; Employee\ Flows_{i,j,t} + Fund\ Characteristics_{i,t} + FEs + \varepsilon_{i,j,t+1} \quad (3)$$

where *i* indexes fund, *j* indexes sponsor fund family, and *t* indexes year. The baseline performance measure is one-year ahead *Carhart Alpha*, estimated using monthly net returns. Standard errors are estimated allowing errors to be correlated within same-fund observations.

Table 2 reports the regression results. When regressing *Carhart Alpha* on Employee Flows without any control variables (Panel A, Column (1)), a one-standard-deviation increase in *Employee Flows* is associated with a significant increase in *Carhart Alpha* of 16bp.

To examine the possibility that the observed predictive power of *Employee Flows* is driven by fund family employees trading into or out of a particular style during a particular year, I estimate the previous regression with fund *style* × *year* fixed effects (Panel A, Column (2)). This ensures that the identification of the effect of *Employee Flows* comes from funds within the same year and style category. The results suggest that my inference is not driven by a particular style-year combination.

The analysis so far implicitly assumes that *Employee Flows* are only driven by the employees' expectation of future performance. Clearly, however, *Employee Flows* will also reflect circumstances specific to a fund family or time period that are unrelated to future fund performance. In general, *Employee Flows* will reflect the constraints imposed on employees by their family and plan at a given point in time. For example, a family with few funds in its plan will tend to have higher *Employee Flows* for a given fund than a family with many funds in its plan (or the same family after it has expanded its menu to include more funds). *Family* × *year* fixed effects effectively control for such non-informative variation in *Employee Flows* across family-years. Thus, one would expect *Employee Flows* to be a stronger predictor of fund performance in the presence of such fixed effects. To examine this conjecture, I estimate a regression with *family* × *year* fixed effects (Panel A, Column (3)). Consistent with the conjecture, the *Employee Flows* coefficient increases by more than 30% relative to the specification without *family* × *year* fixed effects (from 0.192 in Column (1) to 0.253 in Column (3)).

Column (4) of Panel A (Table 2) reports the results for a regression with even more granular *family* × *year* × *style* fixed effects. The identification in this specification comes from the variation in *Employee Flows* relative to funds of the *same* investment style within the *same* fund family' plan in the *same* year. This is the preferred specification as it best captures the ranking of expectation of future performance faced by a fund family's employees.[5] The resulting *Employee Flows* coefficient of 0.240 implies that a one-standard-deviation increase in *Employee Flows* is associated with an increase in one-year ahead *Carhart Alpha* of 21bp. The number of observations

[5] I thus use family × year × style fixed effects in the subsequent analysis. All results, however, are robust to using family × year or style × year fixed effects.

in this specification drops by 20% which reflects the number of family-year-style singleton observations, that is, instances when a family's plan only includes one fund in a given style and year.

To examine whether the relation between *Employee Flows* and future fund performance can be explained by other fund attributes and skill measures that have already been associated with future fund performance in the literature, I consider the following explanatory variables, in addition to *family* × *year* × *style* fixed effects: (i) the natural logarithm of fund size (to control for diminishing returns to scale at the individual fund level), (ii) the natural logarithm of family size (to control for increasing returns to scale at the family level, see, e.g., Chen et al., 2004), (iii) the natural logarithm of one plus fund age (to control for the possibility that older funds, which have survived longer, are better managed), (iv) expense ratios (which are persistent and thus expected to adversely affect future net alpha, see Fama and French, 2010), and (v) past alpha (which could be positively linked to future alpha either through managerial skill, as suggested by Bollen and Busse, 2005) among others, or through flow-induced price pressure, as suggested by Lou, 2012). I also include a range of proxies for skill that have been shown to predict fund performance (though, again, mostly at shorter horizons of 1-3 months): (i) Tracking Error (Cremers and Petajisto, 2009; Grinold and Kahn, 2000), (ii) Return Gap (Kacperczyk, Sialm, and Zheng, 2008), (iii) Active Share (Cremers and Petajisto, 2009), (iv) R^2 (Amihud and Goyenko, 2013), and (v) Abnormal Turnover (Pastor, Taylor, and Stambaugh, 2017).

Column (1) of Panel B (Table 2) shows that controlling for these fund characteristics does not affect the significant relation between *Employee Flows* and future *Carhart Alpha*. Except for abnormal turnover, which is significantly related to future alphas in the expected (positive) direction, none of the other skill measures show predictive power for future *Carhart Alpha* at the one-year horizon.

To examine whether *Employee Flows* contain different information about future fund performance than *Non-Employee Flows* or aggregate flows, I estimate regressions with these other flow measures as additional explanatory variables. Column (2) of Panel B shows that, other things equal, *Non-Employee Flows* are not significantly related to subsequent fund performance. When including both *Employee Flows* and *Non-Employee Flows* as regressors, only the coefficient for *Employee Flows* is significant (see Column (3) of Panel B). Columns (4) and (5) of Panel B report the corresponding regressions with aggregate flows as the additional regressor. Similar to *Non-Employee Flows*, aggregate flows fail to significantly predict one-year ahead *Carhart Alpha* whereas the predictive power of *Employee Flows* is unchanged. In a related analysis, Sialm, Starks, and Zhang (2015) report that DC plan flows – aggregated across all plans in their paper as opposed to flows in one fund family plan in this paper – fail to predict subsequent fund performance. One interpretation for these results, explored in depth in Section III of the paper, is that fund family employees are closer to the fund manager than other fund investors and exploit this proximity to learn about manager skill.

The take-away from Table 2 is that *Employee Flows* have predictive power for one-year ahead *Carhart Alpha* and that this predictive power is essentially orthogonal to known predictors of alpha.

B. The Predictive Power of Employee Flows Is Long-Lived

Existing skill measures and aggregate flows have been reported to predict alphas mostly at short horizons of one to three months; a notable exception is abnormal turnover (Pastor, Taylor, and Stambaugh, 2017). On the one hand, short-term predictability is consistent with a world with decreasing returns to fund scale in which investors redirect capital in response to signals about managerial ability (Berk and Green, 2004). On the other hand, the absence of longer-term predictability gives rise to alternative explanations unrelated to skill. Wermers (2003) and Lou (2012) conjecture that the short-term predictive power of flows, in particular, may be driven by flow-induced price pressure. Consistent with this conjecture, Frazzini and Lamont (2008) and Lou (2012) report that the predictive power of aggregate flows reverses at horizons of two to three years.

To examine whether the predictive power of *Employee Flows* is subject to a similar reversal, I use 2-year- and 3-year ahead *Carhart Alphas* as dependent variables. The use of longer-term alphas leads to overlapping observations and raises concerns about relying on standard test statistics. I address these concerns in one of two ways. First, following Petersen (2009), I double-cluster standard errors on both fund and year. Second, I estimate autocorrelation-and-heteroscedasticity consistent standard errors using a Bartlett kernel (Millo, 2017).

Column (1) of Table 3 reports the results of 2-year-ahead regressions with *family* × *year* × *style* fixed effects. Regardless of the standard error adjustments, *Employee Flows* significantly predict 2-year-ahead *Carhart Alpha*. For example, a one-standard-deviation increase in *Employee Flows* is associated with a 36bp increase in 2-year-ahead *Carhart Alpha*.

Columns (2) of Table 3 reports the corresponding results for 3-year-ahead regressions. Although *Employee Flows* are no longer significant in these regressions at conventional levels, the coefficients have a similar magnitude.

The take-away from Table 3 is that the predictive power of *Employee Flows* for fund performance is long-lived and not subject to the reversals at longer horizons reported for aggregate flows.

C. The Predictive Power of Employee Flows Is Robust

To examine the robustness of the predictive power of *Employee Flows* for mutual fund performance, I consider alternative definitions for *Employee Flows* as well as fund performance.

Like the standard measure of aggregate flows examined in the literature, the baseline definition of *Employee Flows* measures flows in relative terms and implicitly assumes that active changes happen at the end of the calendar year. This definition can also give rise to flows less than -100%, which could happen when investors aggressively sell a strongly performing fund (even if the fund continues to be part of the plan). The first alternative definition measures flows in dollar terms while maintaining the assumption that active changes happen at the end of the calendar year:

$$A1: Employee\ Flows_{i,j,t} \equiv Employee\ Assets_{i,j,t} - Employee\ Assets_{i,j,t-1}\left(1 + r_{i,t}\right) \tag{4}$$

The second alternative definition starts from the baseline measure, but assumes that active changes happen at the beginning of the calendar year (see Zheng 1999).

$$A2: Employee\ Flows_{i,j,t} \equiv \frac{\frac{Employee\ Assets_{i,j,t}}{(1+r_{i,t})} - Employee\ Assets_{i,j,t-1}}{Employee\ Assets_{i,j,t-1}} \tag{5}$$

The third alternative definition takes the simple average of baseline measure and A2, that is, averaging across the measures calculated under the extreme assumptions of active changes happening either at the end or at the beginning of the year.

$$A3: Employee\ Flows_{i,j,t} \equiv \frac{Baseline\ Employee\ Flows_{i,j,t} + A2_{i,j,t}}{2} \tag{6}$$

The fourth alternative definition normalizes the dollar employee flows by the projected value of the lagged plan balance assuming no new investments or redemptions in the fund. This normalization ensures that flows never fall below -100% (see Sialm, Starks, and Zhang, 2015).

$$A4: Employee\ Flows_{i,j,t} \equiv \frac{Employee\ Assets_{i,j,t} - Employee\ Assets_{i,j,t-1}(1+r_{i,t})}{Employee\ Assets_{i,j,t-1}(1+r_{i,t})} \tag{7}$$

Panel A of Table 4 reports the results of regressing one-year-ahead *Carhart Alpha* on the alternative measures of *Employee Flows* as well as control variables and *family* × *year* × *style* fixed effects. The results remain statistically significant and economically meaningful regardless of the definition of *Employee Flows*.

Besides *Carhart Alpha*, I consider four alternative measures of fund performance: (i) Gross *Carhart Alpha* (by adding the expense ratio), (ii) Fama-French 3-factor Alpha (net), (iii) CAPM Alpha (net), and (iv) Raw Return (net). Panel B of Table 4 reports the results of regressing these alternative fund performance measures on baseline *Employee Flows* as well as control variables and *family* × *year* × *style* fixed effects. The results remain statistically significant and economically meaningful regardless of the definition of fund performance.

The overall take-away from Table 4 is that the predictive power of *Employee Flows* for future fund performance is robust to alternative definitions of *Employee Flows* and fund performance.

III. Employees Exploit Proximity to Managers to Learn About Skill

This section explores the mechanism behind the predictive power of *Employee Flows* documented previously. I present evidence consistent with employees exploiting their proximity to fund managers to learn about manager skill (the "information transmission" explanation). If fund family employees exploited their proximity to fund managers to learn about manager skill, one would expect the predictive power of *Employee Flows* to be greater if employees were closer to fund managers. Such an explanation is consistent with both the predictive power of *Employee Flows* documented in this paper and the lack of predictive power of employee flows aggregated across many DC plans reported by Sialm, Starks, and Zhang (2015). Fund family employees are closer to fund managers than employees at other firms in many respects: occupational, organizational, and geographic, to name a few. The following two subsections explore heterogeneity in terms of organizational and geographic distance between fund employees and fund managers in the sample.

A. Employee Flows Predict the Performance of Internal Funds

In principle, a plan fund can be either advised by the family who sponsors the plan ("internal fund") or by a different fund family ("external fund"). I classify a fund as internal only if the plan sponsor family serves as the sole advisor to the fund. The classification is based on advisory details from NSAR filings that mutual funds are mandated to report to the SEC on a semi-annual basis. More than two thirds of fund families include external funds in their plan line-up; external fund-years represent 32% of the total number of fund-years in my sample. In untabulated results, I find that external funds tend to be smaller than internal funds, have slightly higher expense ratios (average of 99bp versus 96bp), and appear to be slightly more active (e.g., Active Share of 84% versus 78%).

Presumably, firm boundaries make it more difficult or costly for fund family employees to gather signals about manager skill for external funds than for internal funds. Thus, if information transmission within a family drove the predictive power of *Employee Flows*, one would expect the predictive power to be greater for the performance of internal funds than that of external funds. To examine this conjecture, I estimate the predictive panel regression separately for the sample of internal funds and for the sample of external funds. This empirical specification reflects the conjecture that fund family employees can discriminate across internal funds, but not across external funds (or across a mix of internal and external funds).

Column (1) of Table 5 reports the regression results for internal funds including controls and *family* × *year* × *style* fixed effects. Column (2) of Table 5 reports the corresponding regression results for external funds. The predictive power of *Employee Flows* is statistically significant and economically meaningful for internal funds. For example, a one-standard-deviation increase in *Employee Flows* is associated with an increase in 1-year-ahead *Carhart Alpha* of internal funds of 26bp. In contrast, the predictive power of *Employee Flows* for external funds is statistically insignificant. The *Employee Flows* coefficients in Columns (1) and (2) are not significantly different from each other, however. It is possible that the lack of significance for external funds is due to the relatively small number of observations.

The take-away from Table 5 is that firm boundaries (organizational distance) between fund employees and managers weaken the predictive power of *Employee Flows*, which is consistent with employees exploiting their proximity to fund managers to learn about manager skill.

B. Employee Flows Predict the Performance of Funds in Single-Location Families

The geographic organization of fund families offers another test of the information transmission explanation. Fund families operate either in a single location or in multiple locations. Operating in a single location should facilitate the information transmission within the organization. Thus, one would expect the predictive power of *Employee Flows* for the performance of internal funds to be stronger for single-location families. To determine whether a fund family operates from a single location or multiple locations, I extract the location of the portfolio managers of the family-advised funds from the funds' semi-annual NSAR filings.

To examine the hypothesis, I estimate the predictive panel regression separately for the sample of internal funds from single-location families and for the sample of internal funds from multiple-location families. Column (1) of Table 6 reports the regression results for funds advised by single-location families, controlling for fund characteristics and *family* × *year* × *style* fixed effects. Column (2) of Table 6 reports the corresponding regression results for funds advised by multiple-location families. The predictive power of *Employee Flows* is statistically significant and economically meaningful for funds advised by single-location families. A one-standard-deviation increase in *Employee Flows* is associated with an increase of one-year-ahead *Carhart Alpha* of 29bp for such funds. In contrast, the predictive power of *Employee Flows* for funds advised by multiple-location families is neither statistically significant nor economically meaningful (though it is not significantly different from the *Employee Flows* coefficient in Column (1), either).

The take-away from Table 6 is that geographic distance between fund employees and managers weakens the predictive power of *Employee Flows*, which again is consistent with employees exploiting their proximity to fund managers to learn about manager skill. The distance-based differences in predictive power reported in Tables 5 and 6 also imply that employee sophistication, by itself, is unlikely to explain the predictive power of *Employee Flows* for fund performance.

The distance-based differences in predictive power are also consistent with proximity breeding effort. Knowing that their fellow employees are betting their retirement money on their funds, fund managers may exert greater effort and deliver better performance. With the data at hand, it is difficult to distinguish between an explanation based on information transmission and effort-based explanation.

IV. Alternative Explanations

The results in the previous sections suggest that predictive power of *Employee Flows* for future fund performance arises due to recognition of managerial skill by the fund family employees. I examine three alternative explanations. First, plan design by fund families may drive the predictive power of *Employee Flows*. Second, it is possible that the predictive power of *Employee Flows* is driven by portfolio managers 'ownership. Third, the predictive power of *Employee Flows* may reflect the employees' advance knowledge of cross-subsidization activities rather than manager skill.

A. Plan Design Is Unlikely to Explain the Predictive Power of Employee Flows

Plan design, understood as the fund family (plan sponsor) adding funds to or dropping funds from the plan, cannot explain the results because *Employee Flows* are constructed to reflect participant actions rather than sponsor actions.

It is still possible that fund families improve their plan menu over time by keeping or adding high-performing funds and dropping low performers. To examine this, I estimate a panel regression of one-year-ahead *Carhart Alpha* on a plan dummy that is equal to 1 if a fund is an In-Plan-Fund and 0 otherwise. If families actively managed their plans towards higher-performing funds, one would expect the dummy coefficient to be significantly positive. Table 7 reports the regression results. Across all fund-years, In-Plan-Funds outperform Not-In-Plan funds by an insignificant 5bp during the subsequent year, not controlling for any fixed effects or fund characteristics. Columns (2)-(5) report the estimation results with various fixed effects and fund characteristics. Depending on the control variables, In-Plan-Funds outperform their peers by up to 26bp per year, but the outperformance is not statistically significant.

In principle, it is still possible that the fund family as plan sponsor knows which plan funds are likely to outperform in the future and implicitly promotes these funds, either by selecting the plan default from this subset of funds, or by making them more salient in the menu design. This conjecture is challenging to test directly, as fund families are not mandated to report detailed information about the choice architecture such as default or preferential options in Form 5500. Some families voluntarily do so and in all of these cases, the plan default is either a target-date fund or an index fund, both of which are excluded from my analysis and thus cannot drive the results. Moreover, the promotion story cannot explain why the predictive power of *Employee Flows* is significant only for single-location families; multi-location families should have similar promotion incentives. Finally, 19 out of 34 fund families have been sued by their employees regarding their 401(k) plan designs during the sample period, which is at odds with fund families actively steering their employees towards higher-performing funds.

Taken together, the evidence suggests that sponsor actions are unlikely to explain the predictive power of *Employee Flows* for fund performance.

B. Managerial Ownership Is Unlikely to Explain the Predictive Power of Employee Flows

Khorana, Servaes and Wedge (2007) report that larger portfolio manager ownership is associated with a higher *Carhart Alpha* in a one-year sample of 1,400 funds. If portfolio managers held and changed the stakes in their funds through 401(k) plans, *Employee Flows* would reflect the managers' investments in their own funds which could explain the predictive power of *Employee Flows*, in principle. This could also explain why *Employee Flows* predict the performance of internal funds, but not of external funds: *Employee Flows* are observed for a particular fund family and do not include the 401(k) investments of external fund managers (because they are employees of a different fund family).

To examine this possibility, I manually extract fund manager ownership for all internal funds each year from the funds' annual Statements of Additional Information (SAI). I limit the sample to internal funds because managerial ownership in external funds cannot drive variation in *Employee Flows* as just noted. Four out of five funds in the sample exhibit managerial ownership. Only two out of five fund families in my sample, however, state that fund managers hold stakes in their funds through their retirement plan. Moreover, there are several instances in which the managerial ownership reported in the SAI exceeds the fund's entire retirement plan balance reported in Form 5500's Schedule of Assets. Thus, one would not necessarily expect managerial ownership in a fund to be highly correlated with fund family retirement plan balances in that fund. Indeed, the correlation is only 0.25.

I use the ownership data to construct two proxies for a fund manager's investments in his own fund: (i) relative ownership defined as dollar ownership at the mid-point of the reported range divided by fund assets (the measure suggested by Khorana, Servaes and Wedge, 2007), and (ii) absolute dollar ownership (the measure suggested by Ibert, 2019). The correlations between these managerial ownership measures and *Employee Flows* are virtually zero (-0.006 and 0.008, respectively). This could be because managerial ownership cannot be precisely estimated due to the coarseness with which the information is reported, or because managers invest in their own funds outside retirement plans, or because managers invest differently in their own funds, reflecting different incentives. For example, managers have a signaling motive to invest in their own funds whereas other employees have no such motive.

Column (1) Table 8 reports the results of predictive panel regressions with relative ownership as additional regressor and *family* × *year* × *style* fixed effects. Consistent with Khorana, Servaes and Wedge (2007), higher relative ownership predicts higher subsequent *Carhart Alpha*; the relation is insignificant though. Importantly, however, the inclusion of relative ownership does not affect the coefficient of *Employee Flows* (0.283 versus 0.282 in Column (1) of Table 5). Column (2) of Table 8 shows the corresponding results when dollar ownership is included as a regressor. Again, the inclusion of the managerial ownership proxy does not affect the predictive power of *Employee Flows*. Theoretically, it is still possible that *Employee Flows* proxy for portfolio manager activity and do so more precisely than proxies constructed from the coarse portfolio manager ownership reports. If that were the case, however, one would expect variation in *Employee Flows* to forecast future fund performance regardless of the distance between fund managers and employees – contrary to the results reported in the previous section.

The key take-away from Table 8 is that portfolio manager ownership is unlikely to drive the predictive power of *Employee Flows* for fund performance.

C. Cross-Subsidization Is Unlikely to Explain the Predictive Power of Employee Flows

Fund families may have an incentive to favor some funds in their line-up at the expense of others. Gaspar, Massa, and Matos (2006) report that high-fee funds receive a disproportionate share of underpriced IPOs. Eisele et al. (2019) report that high-fee funds benefit from advantageous prices in cross-trades with other funds in their family, but that this effect occurs prior to regulatory reforms in 2004, that is, before the sample period considered in this paper. Chaudhuri, Ivković, and Trzcinka (2018) conjecture that a family's cross-subsidization efforts will likely focus on funds that are both strong recent performers and relatively small in their family. Consistent with this conjecture, they report that such funds outperform their benchmarks.

It is possible that *Employee Flows* reflect knowledge of which fund will be cross-subsidized rather than which fund manager is skilled. This alternative explanation could also account for the predictive power of *Employee Flows* being limited to internal funds and funds of single-location families. Presumably, the proximity of employees to portfolio managers and family executives would facilitate the transmission of knowledge about cross-subsidization strategies across employees.

To examine whether cross-subsidization drives the predictive power of *Employee Flows*, I adopt a two-pronged approach. First, I compute an IPO-adjusted *Carhart Alpha* by explicitly estimating the contribution of IPO allocations to fund returns. If the predictive power of *Employee Flows* were due to employees trading into funds that are subsequently cross-subsidized via IPO allocations, *Employee Flows* should not predict IPO-adjusted *Carhart Alpha*. Second, I control for fund attributes that prior work has linked to cross-subsidization such as past performance, expense ratio, fund age, and, in particular, the interaction between strong past performance and small relative size.

To estimate IPO-adjusted *Carhart Alpha*, I follow Gaspar, Massa, and Matos (2006) and estimate the return contribution of an allocation for a given IPO and fund as (the number of IPO shares held by a fund at the end of the quarter during which the IPO occurs) times (the difference between the first-day closing price and the offer price) divided by (the fund's TNA at the end of the month prior to the IPO). For a given fund and month, I sum IPO contributions across all IPOs for that fund and month. I then subtract this sum from the fund's monthly return to get the fund's IPO-adjusted monthly return. The fund's IPO-adjusted *Carhart Alpha* can be estimated from these

IPO-adjusted monthly returns using the standard procedure described in Section I.B2. The average annual contribution of IPO allocations to fund returns is 1bp. The difference between this small contribution and the contribution range of 9bp-47bp reported by Gaspar, Massa, and Matos (2006) is due to their sample period (1992-2001) capturing the IPO boom of the late 1990s whereas my sample period captures the IPO dearth starting in 2008/9. Column (1) of Table 9 reports results of a predictive panel regression with a similar specification as in Table 5, but with IPO-adjusted *Carhart Alpha* as the dependent variable. The predictive power of *Employee Flows* is unaffected. Column (2) of Table 9 reports the results of a regression controlling for the interaction of strong recent performance and small relative size, as defined in Chaudhuri, Ivković, and Trzcinka (2018). The coefficient on the interaction term is positive, consistent with the hypothesized effect of cross-subsidization on fund performance, but not significant in my sample. Importantly, the inclusion of additional controls designed to capture cross-subsidization activities leaves the coefficient on *Employee Flows* virtually unchanged.

The take-away from Table 9 is that the predictive power of *Employee Flows* is unlikely due to employee knowledge of strategic cross-subsidization at the family level.

V. Can Other Investors Benefit By Mimicking *Employee Flows*?

To address the question of whether non-family-employee investors can improve their welfare by conditioning their fund trading decisions on *Employee Flows,* I examine the performance of a trading strategy that buys fund with the highest past *Employee Flows*. An important consideration in designing the trading strategy is the timing of information disclosure. Although fund families have to provide information about the holdings of their own 401(k) plan participants as of the end of each plan year, they are only required to file this information within 10 months from the end of the plan year. During the sample period, most fund families fully exhaust the filing period.

For a given year and plan, I sort mutual funds into quintiles based on their past year *Employee Flows*. Sorting by year and plan reflects that *Employee Flows* may differ across plans for non-informative reasons (which corresponds to adding *family* × *year* fixed effects in the regressions). Sorting by year and plan also guides the choice of quintiles rather than deciles; plans with fewer than ten actively managed U.S. equity funds would not contribute any observations to the top or bottom deciles. Funds in the bottom (1st) quintile typically experience substantial outflows with median *Employee Flows* of -12%. In contrast, funds in the top (5th) quintile experience median *Employee Flows* of 34%.

To account for the ten-month disclosure delay, I buy an equally-weighted portfolio of top quintile funds at the beginning of November of the following year and hold this portfolio for the following 12 months. Using the time-series of monthly returns for the top quintile portfolio, I estimate its Carhart Alpha; I repeat these estimation steps for the other quintile portfolios in terms of *Employee Flows*. Rows (1) to (5) of Table 10 report the Carhart Alphas of the different quintile portfolios along with their factor loadings. Carhart Alpha increases almost monotonically from the bottom quintile to the top quintile portfolio in terms of *Employee Flows.* The top quintile portfolio of funds has an alpha of -0.6%, which is not significantly different from zero. In contrast, the bottom quintile of funds in terms of *Employee Flows*, that is, funds that tend to be aggressively sold by fund family employees, has a significantly negative alpha of -2.2% per year; the alpha difference between the top and bottom quintile portfolio of 1.6% is statistically significant and economically meaningful (Row (6)). To provide broader performance context, Row (7) reports the Carhart Alpha and factor loadings of an equally-weighted portfolio of all funds that appear in fund family 401(k) plans; this portfolio has a Carhart Alpha of -1.4% per year, which is significantly lower than the alpha of the top portfolio in terms of *Employee Flows* (see Row (8)). Row (9) reports the corresponding attributes for an equally-weighted portfolio of all funds offered by the sample fund families between 2009 and 2016, regardless of whether they appear in a family 401(k) plan or not; this portfolio has a Carhart Alpha of -1.5% per year, which is significantly lower than the alpha of the top portfolio in terms of *Employee Flows* (see Row (10)). Put differently, a strategy that invested in the funds with the highest *Employee Flows* would have outperformed a naïve strategy of investing equally across all actively managed funds offered by the sample fund families by 0.9% per year – this outperformance is both statistically significant and economically meaningful. The take-away from Table 10 is that investors can use *Employee Flows* to identify actively managed funds that perform well in the future relative to other actively managed funds.

To put the performance of buying funds with high *Employee Flows* in perspective, I compare it with the performance of univariate trading strategies based on other alpha predictors discussed in Section 2: aggregate flows, past *Carhart Alpha*, expense ratio, tracking error, Return Gap, Active Share, R^2, Abnormal Turnover. In a first comparison, I use a one-year prediction horizon and focus on funds offered in 401(k) plans of fund families in my sample. In contrast to the 10-month disclosure delay for the strategy based on *Employee Flows*, I only impose a one-quarter delay for the Return Gap and no delays for the remaining measures. In a second comparison, I evaluate strategy performance using all the funds offered by the sample fund families and adopt the methodological baseline choices made by the corresponding papers regarding the granularity of the sort (e.g., deciles as opposed to quintiles) and prediction horizon (e.g., quarters as opposed to years).

Columns (1)-(3) of Table 11 report the results for the first comparison. None of the top quintile portfolios in terms of alternative alpha predictors deliver a significant positive one-year ahead alpha during the sample period. In fact, with the exception of the top portfolio based on *Employee Flows*, the alphas of the top portfolios are significantly negative during the sample period. Moreover, only sorts based on *Employee Flows* and expense ratios generate top portfolio performances that are significantly different from the bottom portfolio performances.

Columns (4)-(6) of Table 11 report the corresponding results for the second comparison, which can be viewed as an out-of-sample test for the different measures, because this paper's performance evaluation period does not overlap with the evaluation periods considered in the literature. None of the top portfolios deliver significantly positive Carhart Alpha. In fact, with the exception of the top portfolio in terms of (lowest) expense ratio, all top portfolios deliver significantly negative Carhart Alphas, ranging from -0.97% (aggregate flows) to -2.37% (return gap). Only sorts based on expense ratio and abnormal turnover generate top portfolio performances that are significantly different from the bottom portfolio performances in the expected direction.

The overall take-away from the paper is that *Employee Flows* are informative about managerial skill and that investors can use this information to identify actively managed funds that perform relatively well in the future.

VI. Conclusion

Employee Flows – flows in retirement plans of fund families – are smart. Fund family employees trade out of retirement plan funds that perform poorly in the future and into funds that perform relatively well. The predictive power of employee flows is long-lived and orthogonal to that of other known fund alpha predictors. It is strongest when fund family employees are located close to fund managers, suggesting that employees exploit their proximity to learn about manager skill. The results have implications for other investors who may be able to use the information contained in *Employee Flows* to avoid poorly performing funds and instead choose funds that perform well relative to other actively managed funds. To the extent that fund employees trade on "insider" information about their funds, the results raise the policy question of whether institutional money managers should be required to disclose more information, or more timely information, about the investments of their employees in the family's products.

REFERENCES

Amihud, J. and Goyenko, R., 2013, Mutual fund R^2 as predictor of performance, *Review of Financial Studies* 26, 667–694.

Berk, J. and Green, R.C., 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269–1295.

Berk, J., van Binsbergen, J. H. and Liu, B., 2017, Matching capital and labor. *Journal of Finance* 72, 2467–2503.

Bollen, N. P. and Busse, J. A., 2005, Short-Term Persistence in Mutual Fund Performance, *Review of Financial Studies* 18, 569–97.

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

Chaudhuri, R., Ivković, Z. and Trzcinka, C., 2017, Cross-subsidization in institutional asset management firms, *Review of Financial Studies* 31, 638–677.

Chen, J., Hong, H., Huang, M. and Kubik, J. D., 2004, Does fund size erode mutual fund performance? The role of liquidity and organization, *American Economic Review* 94, 1276–302.

Cremers, K.J. M. and Petajisto, A., 2009, How active is your fund manager? A new measure that predicts performance, *Review of Financial Studies* 22, 3329–3365.

Daniel, N., Dorn, D. and Pedersen, D., 2018, Targeting targets, Working Paper.

Duflo, E. and Saez, E., 2003, The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment, *Quarterly Journal of Economics* 118, 815–42.

Eisele, A., Nefedova, T., Prise, G. and Peijnenburg, K., 2020, Trading out of sight: An analysis of cross-trading in mutual fund families, *Journal of Financial Economics* 135, 359–378

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

Fama, E. F. and French, K.R., 2010, Luck versus skill in the cross-section of mutual fund returns, *Journal of Finance* 65, 1915–47.

Fama, E.F. and French, K.R., 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Frazzini, A. and Lamont, O. A., 2008, Dumb money: Mutual fund flows and the cross-section of stock returns, *Journal of Financial Economics* 88, 299–322.

Gaspar, JM., Massa, M. and Matos, P., 2006, Favoritism in mutual fund families? Evidence of strategic cross-fund subsidization, *Journal of Finance* 61, 73–104.

Grinold, R. C. and Kahn, R.N., 2000, Active portfolio management, 2nd ed. New York: McGraw-Hill.

Gruber, M., 1996, Another puzzle: The growth in actively managed mutual funds, *Journal of Finance* 51, 783–810.

Ibert, M., 2019, What Do Mutual Fund Managers' Private Portfolios Tell Us About Their Skills?, Working Paper.

Kacperczyk, M., Sialm, C. and Zheng, L., 2007, Unobserved actions of mutual funds, *Review of Financial Studies* 21, 2379–2416.

Keswani, A. and Stolin, D., 2008, Which money is smart? Mutual fund buys and sells of individual and institutional investors, *Journal of Finance* 63, 85–118.

Khorana A., Servaes H., Wedge L..2007, Portfolio manager ownership and fund performance, *Journal of Financial Economics*, 85, 179–204.

Levenshtein VI., 1965, Binary codes capable of correcting deletions, insertions, and reversals. Doklady Akademii., 163(4), 845–848.

Lou, D., 2012, A flow-based explanation for return predictability, *Review of Financial Studies* 25, 3457–3489.

Mann, W., 2018, Creditor rights and innovation: evidence from patent collateral, *Journal of Financial Economics* 130, 25–47.

Millo, G., 2017, Robust standard error estimators for panel models: A unifying approach. *Journal of Statistical Software* 82 (1), 1–27.

Ouimet, P. and Tate, G., 2019, Learning from coworkers: Peer effects on individual investment decisions, *Journal of Finance* 0, 1–40.

Pastor, L., Stambaugh, R.B. and Taylor, L., 2017, Do funds make more when they trade more? *Journal of Finance* 72, 1483–1528.

Petersen, M. A., 2009, Estimating standard errors in finance panel data sets: Comparing approaches, *Review of Financial Studies* 22, 435–480.

Sapp, T. and Tiwari, A., 2004, Does Stock Return Momentum Explain the "Smart Money" Effect*?, Journal of Finance 54,* 2605–2622.

Sialm, C., Starks, L. and Zhang, H., 2015, Defined contribution pension plans: Sticky or discerning money? *Journal of Finance* 70, 805–38.

Sirri, E. and Tufano, P., 1998, Costly search and mutual fund flow, *Journal of Finance* 53, 1589–1622.

Wermers, R., 2003, Is money really “smart"? New evidence on the relation between mutual fund flows, manager behavior, and performance persistence, *Working Paper*, University of Maryland.

Zheng, L., 1999, Is money smart? A study of mutual fund investors’ fund selection ability, *Journal of Finance* 54, 901–933.

Table 1: Summary Statistics

The sample comprises actively managed domestic equity funds in the 401(k) plans of the largest U.S. stock fund families between 2009 and 2016. For a given year during the sample period, I identify the largest fund families whose funds collectively account for 80% of the total net assets (TNA) of domestic equity funds in the Center for Research in Security Prices's (CRSP) Survivor-Bias-Free U.S. Mutual Fund Database (MFDB). If a fund family meets the 80% cutoff in any year during the sample period, it is included for the entire sample period; the sample consists of a total of 34 fund families and 336 distinct funds yielding 1,761 fund-year observations. The table reports summary statistics for measures of fund flows and fund alpha. The appendix provides detailed variable definitions for all variables used in the paper.

Fund Flow Measures	Mean	SD
Employee Flows [%]	17.1	82.2
Non-Employee Flows [%]	3.1	44.0
Aggregate Flows [%]	1.5	51.2
Employee Asset / Fund TNA [%]	2.0	4.6
Fund Performance Measures	Mean	SD
12m Carhart Alpha [bp]	-144	433
12m IPO Adjusted Carhart Alpha [bp]	-145	432
24m Carhart Alpha [bp]	-321	702
36m Carhart Alpha [bp]	-557	1003
12m Net Return [%]	12.4	13.5
12m Gross Return [%]	13.5	13.6
Observations	1761	

Table 2: *Employee Flows* Predict Future Alpha

This table summarizes the results of predictive regressions of a fund's one-year-ahead *Carhart Alpha* on *Employee Flows* and fixed effects as well as time-varying fund characteristics. Panel A reports the estimates controlling for different types of fixed effects (FEs). Panel B reports the estimates controlling for Family × Year × Style FEs and time-varying fund characteristics. The sample comprises all actively managed domestic equity funds which are included in the 401(k) plans of the sponsoring fund families. The regression specification is:

$$Carhart\ Alpha_{i,j,t+1} = \gamma\ Employee\ Flows_{i,j,t} + FEs + Fund\ Characteristics_{i,t} + error_{i,j,t+1}$$

where *i* indexes fund, *j* indexes sponsor fund family, and *t* indexes year. All variables are defined in the Appendix. All independent variables are lagged one year. Standard errors are clustered at the fund level and t-statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level respectively.

Panel A:

	(1)	(2)	(3)	(4)
Employee Flows	0.192*	0.233**	0.253**	0.240*
	(1.87)	(2.43)	(2.17)	(1.94)
Style × Year FE		Y		
Family × Year FE			Y	
Family × Year × Style FE				Y
Observations	1761	1761	1744	1429
Adj-Rsq	0.001	0.071	0.110	0.156

Panel B:

	(1)	(2)	(3)	(4)	(5)
Employee Flows	0.269**		0.238**		0.231**
	(2.30)		(2.08)		(2.00)
Non-Employee Flows		0.394	0.328		
		(1.21)	(1.03)		
Aggregate Flows				0.447	0.372
				(1.40)	(1.18)
Carhart Alpha	0.054*	0.049	0.049	0.048	0.048
	(1.67)	(1.51)	(1.49)	(1.47)	(1.46)
R-squared	-0.021	-0.022	-0.020	-0.022	-0.021
	(-0.36)	(-0.37)	(-0.35)	(-0.37)	(-0.35)
Annual return gap	-0.015	-0.016	-0.015	-0.015	-0.015
	(-0.94)	(-0.98)	(-0.93)	(-0.97)	(-0.93)
Active Share	-0.021	-0.020	-0.021	-0.020	-0.021
	(-1.21)	(-1.13)	(-1.18)	(-1.14)	(-1.19)
Abnormal Turnover	0.014*	0.014*	0.014*	0.014*	0.014*
	(1.88)	(1.95)	(1.88)	(1.97)	(1.89)
Tracking Error	0.078	0.066	0.073	0.064	0.071
	(0.31)	(0.26)	(0.29)	(0.25)	(0.28)
Log (Fund Size)	0.001	0.000	0.000	0.000	0.000
	(0.42)	(0.33)	(0.34)	(0.33)	(0.34)
Log (Family Size)	0.002	0.002	0.002	0.002	0.002
	(0.80)	(0.82)	(0.78)	(0.81)	(0.77)
Log (1+Fund Age)	-0.007**	-0.006*	-0.006*	-0.006*	-0.006*
	(-2.12)	(-1.82)	(-1.83)	(-1.80)	(-1.82)
Annual Expense Ratio	-1.561**	-1.464**	-1.544**	-1.458**	-1.536**
	(-2.06)	(-1.97)	(-2.06)	(-1.97)	(-2.05)
Annual Turnover	-0.010***	-0.010***	-0.010***	-0.010***	-0.010***
	(-3.26)	(-3.24)	(-3.29)	(-3.24)	(-3.29)
Family × Year × Style FE	Y	Y	Y	Y	Y
Observations	1429	1429	1429	1429	1429
Adj-Rsq	0.186	0.185	0.186	0.186	0.187

Table 3: *Employee Flows* Predict Alpha Over Longer Horizons

This table summarizes the results of predictive regressions of a fund's *Carhart Alpha* over different horizons on *Employee Flows* and fixed effects as well as time-varying fund characteristics. The sample comprises all actively managed domestic equity funds which are included in the 401(k) plans of sponsor fund families. The regression specification is:

$$Carhart\ Alpha_{i,j,t+h} = \gamma\ Employee\ Flows_{i,j,t} + FEs + Fund\ Characteristics_{i,t} + error_{i,j,t+h}$$

where *i* indexes fund, *j* indexes sponsor fund family, *t* indexes year, and *h* indexes predictability horizon. The dependent variable in Column (1) is 2-year-ahead net *Carhart Alpha* and in Column (2) is 3-year-ahead net *Carhart Alpha*. Fund characteristics are: past alpha, the natural logarithm of fund size, the natural logarithm of family size, the natural logarithm of (1+fund age), turnover, expense ratio, Return Gap, Active Share, R^2, Abnormal Turnover, and Tracking Error. All variables are defined in the Appendix. All independent variables are lagged one year. Reported are the regression coefficients and three sets of t-statistics. The first – in round brackets – are computed using standard errors clustered by fund. The second – in square brackets – are computed using autocorrelation-and-heteroscedasticity consistent standard errors. The third – in curly brackets – are computed using standard errors clustered by fund and by year. ***, **, and * indicate significance at the 1%, 5%, and 10% level respectively.

Horizon →	(1) 2yrs	(2) 3yrs
Employee Flows	0.420**	0.445
	(2.08)	(1.31)
	[2.02]	[1.59]
	{2.95}	{0.91}
Family × Year × Style FE	Y	Y
Fund-Characteristics	Y	Y
Observations	1395	1182
Adj-Rsq	0.177	0.181

Table 4: The Predictive Power of *Employee Flows* Is Robust

This table reports the results of predictive regressions of future fund performance on *Employee Flows* and fixed effects as well as time-varying fund characteristics. The sample comprises all actively managed domestic equity funds which are included in the 401(k) plans of sponsor fund families. All variables are defined in the Appendix. All independent variables are lagged one year. Fund characteristics are the same as in Table 3. Panel A reports the results using alternative measures of *Employee Flows* with one-year-ahead *Carhart Alpha* as the dependent variable. Panel B reports the results using the baseline definition of *Employee Flows* and alternative measures of one-year-ahead performance. Fund characteristics in Column (4) in Panel B excludes the active skill measures and past alpha and add past net return as an additional control. Standard errors are clustered at the fund level and t-statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level respectively.

The alternate measures of Employee Flows are defined as under.

$$A1: Employee\ Flows_{i,j,t} \equiv Employee\ Assets_{i,j,t} - Employee\ Assets_{i,j,t-1}(1 + r_{i,t})$$

$$A2: Employee\ Flows_{i,j,t} \equiv \frac{\frac{Employee\ Assets_{i,j,t}}{(1 + r_{i,t})} - Employee\ Assets_{i,j,t-1}}{Employee\ Assets_{i,j,t-1}(1 + r_{i,t})}$$

$$A3: Employee\ Flows_{i,j,t} \equiv \frac{Baseline\ Employee\ Flows_{i,j,t} + A2_{i,j,t}}{2}$$

$$A4: Employee\ Flows_{i,j,t} \equiv \frac{Employee\ Assets_{i,j,t} - Employee\ Assets_{i,j,t-1}(1 + r_{i,t})}{Employee\ Assets_{i,j,t-1}(1 + r_{i,t})}$$

Panel A: Alternate measures of *Employee Flows*

	(1)	(2)	(3)	(4)
Employee Flows Measure →	*A1*	*A2*	*A3*	*A4*
Employee Flows	0.074*	0.295**	0.283**	0.284**
	(1.84)	(2.23)	(2.27)	(2.27)
Family × Year × Style FE	Y	Y	Y	Y
Fund-Characteristics	Y	Y	Y	Y
Observations	1429	1429	1429	1429
Adj-Rsq	0.186	0.186	0.186	0.186

Panel B: Alternate measures of one-year ahead fund performance

	(1)	(2)	(3)	(4)
Performance Measure →	Gross Carhart	FF3	CAPM	Net Return
Employee Flows	0.269**	0.276**	0.444**	0.340**
	(2.31)	(2.38)	(2.38)	(2.17)
Family × Year × Style FE	Y	Y	Y	Y
Fund-Characteristics	Y	Y	Y	Y
Observations	1429	1429	1429	1429
Adj-Rsq	0.173	0.197	0.324	0.844

Table 5: *Employee Flows* Predict the Performance of Internal Funds

This table summarizes the results of predictive regressions of a fund's one-year-ahead *Carhart Alpha* on *Employee Flows* and fixed effects as well as time-varying fund characteristics.

$$Carhart\ Alpha_{i,j,t+1} = \gamma\ Employee\ Flows_{i,j,t} + FEs + Fund\ Characteristics_{i,t} + error_{i,j,t+1}$$

where *i* indexes fund, *j* indexes sponsor fund family, and *t* indexes year. Column (1) reports results for the sample of internal funds (funds solely advised by the fund family who sponsors the plan) whereas Column (2) reports results for the sample of external funds (funds advised or subadvised by fund families other than the plan sponsor). Fund characteristics are the same as in Table 3. All variables are defined in the Appendix. All independent variables are lagged one year. Standard errors are clustered at the fund level and t-statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level respectively.

	(1)	(2)
Employee Flows	0.282**	0.536
	(2.13)	(1.24)
Family × Year × Style FE	Y	Y
Fund-Characteristics	Y	Y
Observations	963	383
Adj-Rsq	0.238	0.157

Table 6: *Employee Flows* Predict the Performance of Funds in Single-Location Families

This table summarizes the results of predictive regressions of a fund's one-year-ahead *Carhart Alpha* on *Employee Flows* and fixed effects as well as time-varying fund characteristics.

$$Carhart\ Alpha_{i,j,t+1} = \gamma\ Employee\ Flows_{i,j,t} + FEs + Fund\ Characteristics_{i,t} + error_{i,j,t+1}$$

where *i* indexes fund, *j* indexes sponsor fund family, and *t* indexes year. The sample comprises all actively managed domestic equity funds which are included in the 401(k) plans of sponsor fund families and are solely advised by the sponsoring fund family (internal funds). Column (1) reports the results for funds advised by families whose managers operate in a single location; Column (2) reports the results for multi-location families. Fund characteristics are the same as in Table 3. All variables are defined in the Appendix. All independent variables are lagged one year. Standard errors are clustered at the fund level and t-statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level respectively.

	(1)	(2)
Employee Flows	0.281**	-0.031
	(2.10)	(-0.03)
Family × Year × Style FE	Y	Y
Fund-Characteristics	Y	Y
Observations	806	157
Adj-Rsq	0.255	0.111

Table 7: Plan Design Is Unlikely to Explain the Predictive Power of *Employee Flows*

This table summarizes the results of regressing a fund's one-year-ahead *Carhart Alpha* on *Plan dummy* and fixed effects as well as time-varying fund controls.

$$Performance_{i,j,t+1} = \gamma \ Plan\ dummy_{i,j,t} + FEs + Controls_{i,t} + error_{i,j,t+1}$$

where i indexes fund, j indexes fund family, and t indexes year. *Plan dummy* equals to 1 if fund i is part of 401(k) plan of family j in year t, else it is 0. The coefficient on *Plan dummy* is reported in bp. The sample consists of In-Plan Funds (active funds included in the 401(k) plan of a family in the sample) and Not-In-Plan Funds (active funds managed by the family that sponsors the plan, but not included in the 401(k) plan of a family in the sample). Controls are: Past alpha, the natural logarithm of Fund Size, the natural logarithm of (1+Fund Age), turnover, and expense ratio. All variables are defined in the Appendix. All independent variables are lagged one year. Standard errors are clustered at the fund level and t-statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level respectively.

	(1)	(2)	(3)	(4)	(5)
Plan dummy (coeff in bp)	4.842	20.021	17.111	26.126	22.272
	(0.30)	(1.28)	(1.11)	(1.49)	(1.16)
Style × Year FE			Y		
Family × Year FE				Y	
Family × Year × Style FE					Y
Controls		Y	Y	Y	Y
Observations	3711	3711	3711	3698	3437
Adj-Rsq	-0.000	0.019	0.101	0.138	0.225

Table 8: Managerial Ownership Is Unlikely to Explain the Predictive Power of *Employee Flows*

This table summarizes the results of predictive regressions of a fund's one-year-ahead *Carhart Alpha* on *Employee Flows* and fixed effects as well as time-varying fund characteristics.

$$Carhart\ Alpha_{i,j,t+1} = \gamma\ Employee\ Flows_{i,j,t} + FEs + Fund\ Characteristics_{i,t} + error_{i,j,t+1}$$

where *i* indexes fund, *j* indexes sponsor fund family, and *t* indexes year. The sample comprises all actively managed domestic equity funds which are included in the 401(k) plans of sponsor fund families and are solely advised by the sponsoring family (internal funds). Following Khorana, Servaes and Wedge (2007)), dollar ownership is defined as the mid-point of the reported range of managerial ownership divided by fund TNA. Following Ibert (2019), dollar ownership is defined as the dollar value of managerial ownership estimated as the mid-point of the reported range of managerial ownership. Fund characteristics are the same as in Table 3. All variables are defined in the Appendix. All independent variables are lagged one year. Standard errors are clustered at the fund level and t-statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level respectively.

	(1)	(2)
Employee Flows	0.283**	0.282**
	(2.14)	(2.13)
$ Ownership / Fund TNA	0.174	
	(0.48)	
$ Ownership		-0.000
		(-0.24)
Family × Year × Style FE	Y	Y
Fund-Characteristics	Y	Y
Observations	963	963
Adj-Rsq	0.237	0.237

Table 9: Cross-Subsidization Is Unlikely to Explain the Predictive Power of *Employee Flows*

This table summarizes the results of predictive regressions of a fund's one-year-ahead fund performance on *Employee Flows* and fixed effects as well as time-varying fund characteristics.

$$Fund\ Performance_{i,j,t+1} = \gamma\ Employee\ Flows_{i,j,t} + FEs + Fund\ Characteristics_{i,t} + error_{i,j,t+1}$$

where *i* indexes fund, *j* indexes sponsor fund family, and *t* indexes year. The sample comprises all actively managed domestic equity funds which are included in the 401(k) plans of sponsor fund families. The dependent variable in Column (1) is one-year-ahead IPO-Adjusted *Carhart Alpha* and in Column (2) is one-year-ahead *Carhart Alpha*. IPO-Adjusted *Carhart Alpha* is estimated using the adjusted monthly fund returns obtained by subtracting the fund returns due to IPO allocations from the monthly fund returns. The "RelSmall" variable in Column (2) is estimated as in Chaudhuri, Ivković, and Trzcinka (2018): an indicator variable that equals one if the ratio of fund size to average fund size in the family in year *t* is in the bottom quartile, and zero otherwise. "Top-Alpha" is an indicator variable that equals one if the fund's *Carhart Alpha* in year *t* is in the top quintile in its style, and zero otherwise. Fund characteristics are the same as in Table 3. All variables are defined in the Appendix. All independent variables are lagged one year. Standard errors are clustered at the fund level and t-statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level respectively.

	(1)	(2)
Employee Flows	0.269**	0.297**
	(2.31)	(2.56)
RelSmall x Top-Alpha		0.011
		(1.27)
Family × Year × Style FE	Y	Y
Fund-Characteristics	Y	Y
Observations	1429	1429
Adj-Rsq	0.186	0.200

Table 10: Investors Can Benefit by Mimicking *Employee Flows*

The table reports factor loadings from a Carhart (1997) four-factor model and Carhart Alpha for portfolios of funds sorted by *Employee Flows*. At the beginning of each year, funds are sorted into quintiles by their *Employee Flows* during the prior year within each sponsor fund family. I form equally-weighted portfolios of funds within a given quintile and hold them for 12 months starting in November of the following year (the 10-month delay between sort and portfolio formation reflects the delay with which Employee Flows are disclosed to the public). The first five rows reports *Employee Flows* quintile portfolio loadings and alpha. "Top – Bottom" is a hypothetical portfolio that is long the top quintile and short the bottom quintile portfolio. "In-Plan" is an equally-weighted portfolio of all funds offered by the sample families in their 401(k) plans in a given year. "All" is an equally-weighted portfolio of all funds offered by the families regardless of their availability in family retirement plans. The table reports annualized Carhart net alpha as a percentage and the corresponding t-statistics (in parentheses). ***, **, and * indicate significance at the 1%, 5%, and 10% level respectively.

Row	*Employee Flows Quintile*	Alpha (%)	MKT-RF	SMB	HML	UMD
(1)	Bottom	-2.172***	1.014***	0.237***	-0.054**	-0.016
		(-3.57)	(60.70)	(11.37)	(-2.22)	(-0.89)
(2)	Q2	-1.692***	0.991***	0.228***	-0.015	-0.002
		(-3.30)	(70.40)	(12.97)	(-0.75)	(-0.10)
(3)	Q3	-1.075**	0.982***	0.178***	-0.032	-0.016
		(-2.07)	(68.98)	(10.02)	(-1.56)	(-1.04)
(4)	Q4	-1.248**	0.954***	0.156***	-0.078***	-0.038**
		(-1.97)	(55.02)	(7.22)	(-3.09)	(-2.09)
(5)	Top	-0.607	0.945***	0.133***	-0.059**	-0.039**
		(-1.01)	(57.57)	(6.50)	(-2.47)	(-2.25)
(6)	Top - Bottom	1.579***	-0.070***	-0.104***	-0.005	-0.023
		(3.35)	(-5.03)	(-5.73)	(-0.23)	(-1.37)
(7)	In-Plan	-1.428***	0.981***	0.191***	-0.047**	-0.021
		(-2.87)	(71.54)	(11.15)	(-2.38)	(-1.47)
(8)	Top – In-Plan	0.816**	-0.035***	-0.057***	-0.011	-0.017*
		(2.55)	(-4.05)	(-5.22)	(-0.89)	(-1.89)
(9)	All	-1.476***	981***	0.181***	-0.023	-0.019
		(-3.36)	(81.57)	(11.08)	(-1.33)	(-1.50)
(10)	Top – All	0.864**	-0.036***	-0.047***	-0.035***	-0.020**
		(2.61)	(-3.99)	(-4.21)	(-2.70)	(-2.08)

Table 11: Comparison of Univariate Trading Strategies

The table compares the results of different univariate trading strategies from 2010 to 2018. The sample underlying Columns (1)-(3) are funds offered by the sample fund families in their 401(k) plans between 2009 and 2016 ("In-Plan"). At the beginning of each year, funds are sorted by a given strategy variable and quintile portfolios are formed by equally-weighting funds in a given quintile. The sort order is chosen such that the "Top" quintile portfolio would be expected to have better performance, e.g., descending sort for expense ratio and ascending sort for Active Share. The corresponding portfolios are bought either with a 10-month disclosure delay (for *Employee Flows*), a 3-month disclosure delay (for Return Gap) or no disclosure delay (for the remaining strategy variables) and held for 12 months. The sample underlying Columns (4)-(6) are all funds offered by the sample fund families between 2009 and 2016, regardless of their availability in family 401(k) plans. For the trading strategy results reported in Column (4)-(6), I adopt the methodological choices of the corresponding papers regarding granularity of the sort (e.g., quintiles for Active Share and deciles for Return Gap) and the holding period (e.g., 1-month for R-squared, 3-month for aggregate flows). Again, the sort order is chosen such that the "Top" portfolio would be expected to have better performance. The table reports annualized Carhart net alpha as a percentage and the corresponding t-statistics (in parentheses). ***, **, and * indicate significance at the 1%, 5%, and 10% level respectively.

	In-Plan			All		
Predictor	(1) Top	(2) Bottom	(3) Top – Bottom	(4) Top	(5) Bottom	(6) Top – Bottom
Employee Flows	-0.593 (-1.01)	-2.172*** (-3.57)	1.579*** (3.35)			
Active Share	-1.901*** (-2.73)	-1.060** (-2.68)	-0.841 (-1.35)	-1.692** (-2.19)	-1.091*** (-3.77)	-0.610 (-0.84)
Return Gap	-1.753** (-2.19)	-2.051*** (-3.37)	0.298 (0.37)	-2.371** (-2.51)	-2.060*** (-3.80)	-0.311 (-0.34)
$1\text{-}R^2$	-1.665* (-1.92)	-1.721*** (-3.47)	0.056 (0.07)	-1.981*** (-2.80)	-0.711** (-2.14)	-1.270* (-1.78)
Tracking Error	-2.188** (-2.42)	-1.637*** (-4.05)	-0.551 (-0.66)	-1.992** (-2.54)	-1.051*** (-4.01)	-0.941 (-1.23)
Abnormal Turnover	-2.003*** (-2.73)	-2.401*** (-3.83)	0.402 (0.67)	-1.095** (-2.10)	-2.403*** (-4.28)	1.312*** (3.16)
Expense Ratio	-0.995** (-2.40)	-2.924*** (-4.29)	1.931*** (3.65)	0.061 (0.24)	-2.292*** (-3.51)	2.231*** (3.26)
Aggregate Flows	-1.514*** (-3.34)	-1.654*** (-2.70)	0.140 (0.22)	-0.971** (-2.23)	-1.292** (-2.27)	0.321 (0.54)
Past Alpha	-2.255*** (-3.27)	-2.053** (-2.43)	-0.202 (-0.23)	-1.701** (-2.40)	-1.781*** (-2.75)	0.080 (0.07)

Appendix: Variable Definitions

Variable	Definition
Carhart Alpha	Carhart Alpha is estimated using a rolling window method. Specifically, factor betas in month *t* are estimated by regressing monthly fund returns in excess of the one-month T-bill rate on monthly factor realizations during the period *t-1* and *t-36*. Funds need to have a minimum of 24 months of data for the factor beta estimates (and consequently, for the alpha estimates) to be valid. The estimated factor betas are multiplied by the corresponding factor realizations in month *t* to compute a predicted excess return for month *t*. Predicted excess returns and realized excess returns are compounded over a year (or longer periods). *Carhart Alpha* is the difference between the compounded realized excess return and the compounded predicted excess return.
FF3 Alpha	Defined analogously to Carhart Alpha, but by using Fama-French Three-Factor model (1993) to estimate factor betas.
CAPM Alpha	Defined analogously to Carhart Alpha, but by using CAPM to estimate factor beta.
Gross Carhart Alpha	Defined analogously to Carhart Alpha using gross monthly returns.
IPO-Adjusted Carhart Alpha	Defined analogously to Carhart Alpha using IPO-adjusted fund returns. IPO-adjusted return for a fund is defined as the difference between the fund's return and return of IPO allocations to the fund.
Employee Flows	$Employee\ Flows_{i,j,t} \equiv \frac{Employee\ Assets_{i,j,t} - Employee\ Assets_{i,j,t-1}(1+r_{i,t})}{Employee\ Assets_{i,j,t-1}}$ $subject\ to\ Employee\ Assets_{i,j,t-1} > 0$ and $Employee\ Assets_{i,j,t} > 0$ where *i* indexes fund, *j* indexes sponsor fund family, and *t* indexes year. $Employee\ Assets_{i,j,t}$ is the dollar market value of plan balance of fund *i* in the plan of fund family *j* at the end of year *t*. $r_{i,t}$ is return of fund *i* in year *t*.
Non-Employee Flows	Defined analogously to Employee Flows, but using Non-Employee Assets, that is, the difference between a fund's Total Net Assets (TNA) and its Employee Assets.
Aggregate Flows	The net growth in fund assets beyond reinvested dividends over the past one year (see, e.g., Sirri and Tufano, 1998).
Fund Age	Length of time, in years, since the fund inception date (based off the oldest share class).
Annual Expense Ratio	Weighted average of expense ratios of all share classes of a fund using total net assets of each share class as weights.
Fund Size	Total net assets of all share classes of a fund at the end of each year.
Family Size	Total net assets of all actively managed domestic U.S. equity funds by the family at the end of each year.
Annual Turnover	Minimum of aggregate purchases and sales of securities divided by average TNA over the calendar year from the CRSP MFDB.
Active Share	Following Cremers and Petajisto (2009), active share is estimated as one half of the sum of absolute value of the deviation between the weight a fund places on a stock and the weight of the stock in a benchmark index. For a fund, the minimum active share across all benchmark indices is used as the fund's Active Share. I consider the following benchmarks: Russell 1000

	(Composite, Growth, and Value), Russell 2000 (Composite, Growth, and Value), S&P500 (Composite, Growth, and Value), and S&P400.
Annual Return Gap	Following Kacperczyk et al. (2008), Return Gap is estimated as the difference between a fund's gross return and holding return based on a hypothetical portfolio that invest in fund's most recent disclosed holdings.
Tracking Error	Following Cremers and Petajisto (2009), Tracking Error is estimated as the annualized standard deviation of residuals from the annual four-factor model regression of a fund's gross return.
Abnormal Turnover	Abnormal turnover is based on Pastor et al. (2017). Following Daniel, Dorn, and Pedersen (2018), Abnormal Turnover is estimated as the difference between a fund's annual turnover and average turnover of the fund over the sample period.
R-squared	Following Amihud and Goyenko (2013), R-squared is estimated as the R^2 from the four-factor model regression of returns.
$ Ownership	Absolute dollar ownership in a fund is the sum total of dollar ownership across all managers of the fund. Dollar ownership of each manager is estimated as the mid-point of the reported ownership range in the fund's Statement of Additional Information.
$ Ownership / TNA	Following Khorana, Servaes and Wedge (2007), relative ownership is defined as the fund managers' dollar ownership at the mid-point of the reported ownership range, divided by the fund's total net assets.