Smart Fund Managers? Stupid Money?*

Dan Bernhardt†
University of Illinois

Ryan J. Davies‡
Babson College

August 5, 2008

Abstract

We develop a model of mutual fund manager investment decisions near the end of quarters. We show that when investors reward better performing funds with higher cash flows, near quarter-ends a mutual fund manager has an incentive to distort new investment toward stocks in which his fund holds a large existing position. The short-term price impact of these trades increase the fund’s reported returns. Higher returns are rewarded by greater subsequent fund inflows which, in turn, allow for more investment distortion the next quarter. Because the price impact of trades is short-term, each subsequent quarter begins with a larger return deficit. Eventually, the deficit cannot be overcome. Thus, our model leads to the empirically observed short-run persistence and long-run reversal in fund performance. In doing so, our model provides a consistent explanation of many other seemingly contradictory empirical features of mutual fund performance.

JEL Classification: D82, G2, G14.

Keywords: turn-of-quarter effect, painting the tape, mutual fund performance, investment distortion.

---

*Richard Martin’s master’s essay under the direction of the first author was a building point for this research. We greatly appreciate Harvey Westbrook Jr.’s significant contributions to early drafts. We are also grateful for comments received from two anonymous referees, Chris Brooks, Murillo Campello, Duane Seppi, and seminar participants at Queen’s University, St. Francis Xavier University, the Northern Finance Association meetings, the Financial Management Association European meetings, and the Multinational Finance Society meetings.

†Department of Economics, University of Illinois, 1206 South Sixth Street, Champaign IL 61820. Tel: 217–244–5708. email: danber@uiuc.edu

‡Finance Division, Babson College, Tomasso Hall, Babson Park MA 02457–0310. Tel: 781–239–5345. Fax: 781–239–5004. email: rdavies@babson.edu (Corresponding author)
Smart Fund Managers? Stupid Money?

We develop a model of mutual fund manager investment decisions near the end of quarters. We show that when investors reward better performing funds with higher cash flows, near quarter-ends a mutual fund manager has an incentive to distort new investment toward stocks in which his fund holds a large existing position. The short-term price impact of these trades increase the fund’s reported returns. Higher returns are rewarded by greater subsequent fund inflows which, in turn, allow for more investment distortion the next quarter. Because the price impact of trades is short-term, each subsequent quarter begins with a larger return deficit. Eventually, the deficit cannot be overcome. Thus, our model leads to the empirically observed short-run persistence and long-run reversal in fund performance. In doing so, our model provides a consistent explanation of many other seemingly contradictory empirical features of mutual fund performance.

**JEL Classification:** D82, G2, G14.

**Keywords:** turn-of-quarter effect, painting the tape, mutual fund performance, investment distortion.
1 Introduction

In this paper, we develop a model of mutual fund manager investment decisions near the end of quarters. Our basic intuition can be summarized as follows. We begin with the well-established observation that investors reward better performing mutual funds with greater cash inflows. Small improvements in relative performance can generate large increases in future fund cash inflows. Thus, mutual fund managers, whose compensation rises with assets under management, have strong incentives to increase the reported quarterly returns of their funds. We show that one method for doing so is for the fund to distort its end-of-quarter purchases toward stocks in which the fund already holds large positions. The short-term price impact of these trades increases the reported value of the fund’s existing positions, thereby raising the fund’s end-of-quarter reported returns.

A fund’s ability to engage in this distortive end-of-quarter trading is limited by the amount of new cash it has available to invest. Funds with better past performance receive larger cash inflows and this facilitates more aggressive end-of-quarter trading. Of course, this distortion comes at a cost — there is no free lunch. Specifically, the price impacts of these trades eventually decay, decreasing the next quarter’s return, all else equal. However, over time, to overcome the hangover from past distortive trading, funds must engage in ever more distortive trades. Eventually the accumulated hangover proves too much, leading to the long-run reversals in fund performance found in the data.

Thus, our model provides a framework for explaining the puzzling empirical evidence that mutual fund returns exhibit short-run persistence (“hot and cold hands”) and long-run reversals (past top performing funds become future under-performers, even after incorporating their initially superior performance). In addition, the cost of the investment distortions that we model can help explain why mutual funds tend to under-perform non-managed indexes, despite evidence that mutual fund managers have some stock picking ability.

A key component of our model is that the price impact of trades decays over time.

---

1 For example, see Chevalier and Ellison (1997), Ippolito (1992), and Sirri and Tufano (1998).
3 Wermers, Chen, and Jegadeesh (2000) and Pinnuck (2003) observe that stocks purchased by mutual funds tend to outperform stocks that they sell; yet, on average, actively-managed funds under-perform index funds.
Early in a quarter, fund managers may trade so as to minimize investment distortions/price impacts. Later in the quarter, however, enough of the price impact persists through the end of the quarter to make it attractive to distort investments toward assets in place. An extreme manifestation of this is the practice of “high closing”—artificially boosting the price of a stock by purchasing shares just before the market closes on the last day of the quarter.\(^4\)

Trades in less liquid stocks generally have greater price impacts. Thus, our model predicts that managers of funds that specialize in less liquid stocks have greater incentives to distort end-of-quarter investment towards existing holdings. Over time, this behavior will lead these funds to hold undiversified portfolios. This is consistent with the observed high return volatility of specialized funds. Our model’s link between short-run fund performance and the concentration of fund portfolios can help reconcile the findings of Kacperczyk, Sialm, and Zheng (2005) on the performance impact of industry concentration in fund holdings.

Our model suggests that funds may adapt their overall holdings to exploit the incentives modeled here. For instance, smaller funds may focus more on smaller, illiquid stocks. In this way, a small fund can be a “big fish in a small pond.” While the typical trade size of a small fund is too small to have much influence on the price of a large stock, the same trade size can have a large impact on prices of less liquid stocks. Similarly, our model predicts that since younger funds have a more sensitive performance-flow relationship, managers of younger funds have greater incentives to distort investment toward assets in place.\(^5\) Thus, our model predictions are consistent with the empirical findings of Zheng (1999) and Wermers (2003) that smaller funds exhibit greater short-run persistence in performance and even more dramatic long-run reversals in performance.

Our model also suggests that mutual funds will tend to be momentum traders—when a mutual fund’s past superior performance is generated by large positions in well-performing stocks, our model predicts that the mutual fund will continue to increases its positions in those stocks, endogenously generating momentum through the resulting price impact. Our model also predicts that there will be a long-run reversal in performance, i.e., that those subsequent purchases will be earn low long-run returns. San (2007) documents each of these

\(^4\)John Gilfoyle reports that “Nearly everyone seems to agree that high closing is common.” Macleans magazine, July 10, 2000 p39. For example, the Ontario Securities Commission in its case against RT Capital Management, Inc. cites the end of year purchase of 1900 shares of Dia Met at a $.50 premium. A Globe and Mail July 7, 2000 investigation found that in mutual funds on the final trading day of the year, “last-minute leaps beyond the normal market trends strongly suggested a round of ‘portfolio pumping’.”

\(^5\)Chevalier and Ellison (1997) find that younger funds must perform better to attract investment.
empirical regularities for trading by institutions.

Our model takes investors’ decisions to reward higher short-term returns with larger cash inflows as given. There is ample empirical evidence documenting this investor behavior. Our analysis suggests that the best response of investors is not given by this stimulus-response setup—hence the reference to “stupid money” in the title; instead, investors optimal investment decision would need to reflect the strategic incentives of the fund managers. To endogenize the response of individual investors, we would need to extend our model to include heterogeneity in fund manager ability, and integrate learning by investors from mutual fund returns. Bernhardt and Nosal (2008) do this in a strategically simpler setting where a hedge fund manager can augment fund payoffs with zero NPV gambles of his own design. In our strategically richer setting, it is infeasible to endogenize both investor and fund manager behavior, and the reader should be alert to the possibility that conclusions about “stupid money” might be changed in such a setting.

The paper proceeds as follows. The next section provides an overview of related literature. Section 3 sets up the economic environment. Section 4 analyzes how a fund’s existing stock holdings influence its manager’s investment decisions, and derive the consequences of these decisions for persistence in fund returns. Section 5 provides numerical examples. Section 6 concludes. Proofs are in an appendix.

2 Related Literature

Gallagher et al. (2007) use daily trading data of investment managers to provide direct evidence of “painting the tape” and show that gaming behavior is more likely to occur in smaller stocks, growth stocks, and stocks in which the fund has a larger position.

Carhart et al. (2002) provides further extensive empirical support. They find that funds earn tremendous short-run returns near the end of an evaluation period, when trading behavior has the greatest impact on performance: 80% of funds beat the S&P 500 Index on the last trading day of the year (62% for other quarter-end dates), but only 37% (40% other quarters) do so on the first trading day of a new year. This effect is even greater for small-cap funds, which trade less liquid stocks: 91% of small-cap funds beat the index at year’s end, and 70% for other quarter-end dates versus 34% for first quarter trading day. Carhart et al. (2002) reject benchmark-beating hypotheses in favor of strategic behavior similar to that
motivated here. Consistent with our theory, they find funds that performed better in the past year earned 42 basis points higher returns on the last trading day and 29 basis points lower on the first trading day than funds with worse historical performance. In sum, the data strongly support the hypothesis that fund managers respond to the short-run trading incentives that we model, and that they are collectively substantial enough to alter aggregate market outcomes.

Indeed, Bernhardt and Davies (2005) find that strategic trading by fund managers appears to impact returns on aggregate market indexes, so that measuring the impact relative to the index, as Carhart et al. (2002) do, understates the total effect. Specifically, daily returns of the equally-weighted index on the last trading day of a quarter greatly exceed the daily returns on the first trading day of the succeeding quarter, and this return difference rises with the share of total equity held by mutual funds.

Zheng (1999) finds that funds with relatively high returns in one quarter have significantly greater returns than the average mutual fund in the next quarter; and relatively worse performers in one quarter generate lower returns than the average mutual fund the next period. Zheng (1999) shows that the performance persistence is short-lived: more than one quarter into the future, funds that did better in the past under-perform relative to the average fund while historically poor-performing funds do better. This reversal in performance is so strong that cumulative returns of historically-better performers fall below those of worse performers within 30 months.

At the stock-holdings level, Wermers (2003) finds that funds ranked in the top quintile in the prior year ("past winners") beat the S&P 500 by two percent in the next year (after accounting for trading costs and expenses). He also finds that past winners beat past losers (funds in the prior year’s bottom quintile) by five percent in the next year. Wermers (2003) provides empirical support for our theoretical connection between fund flows and trades, showing that flow-related buying drives stock prices up, and that this flow-related buying plays a central role in driving the persistence in fund performance.

In an insightful paper, Berk and Green (2004) develop a model in which rational investors learn from a fund’s performance about its manager’s ability, and allocate funds accordingly. The Berk-Green model can explain both why money flows into mutual funds with recent better performance, and away from those with worse past performance. However, the Berk-Green model also implies that past returns should not predict future performance; which
is inconsistent with short-term performance persistence and long-run performance reversals found in the data. Primary contributions of our model are to reconcile these empirical observations, and to provide explanations for several features of the mutual fund market not explained by the Berk-Green model, such as why stocks purchased by mutual funds outperform those that they sell, yet, on average, mutual funds underperform the market; and the end-of-quarter return patterns and strategic behavior documented by Gallagher et al. (2007), Carhart et al. (2002) and Bernhardt and Davies (2005).

3 The Model

3.1 Model Details

Consider a risk neutral mutual fund manager who can invest in three assets: stock $A$, stock $B$ and cash. The fund manager enters quarter $t$ with an existing position of $S_{At}$ shares in stock $A$ and $S_{Bt}$ shares in stock $B$. The associated share prices are $P_{At}$ and $P_{Bt}$. Without loss of generality we assume that $P_{At}S_{At} \geq P_{Bt}S_{Bt}$. Cash earns a risk-free return of $i$, which we normalize to zero.

Firms retain their earnings. Hence, a firm’s stock price is equal to its expected discounted lifetime cash flows. We impose no structure on the timing of cash flows. At the beginning of quarter $t$, each firm makes an earnings announcement and provides guidance about future cash flows, which causes market participants to update about firm values. We summarize the earnings announcement and guidance by its implications for the percentage change in firm value, $\delta_{j,t}$, so that $\delta_{j,t}P_{j,t-1}$ is the change in expected discounted cash flows. Following a signal of $\delta_{j,t}$, investors update about cash flows, leading to a quarter $t$ stock price for firm $j$ of

$$P_{j,t} = P_{j,t-1} + \delta_{j,t}P_{j,t-1}. \quad (1)$$

We assume that in quarter $t$, the fund manager privately learns the next quarter’s signals, $\delta_{A,t+1}$ and $\delta_{B,t+1}$, and can trade based on this private information in quarter $t$ before the signals are publicly revealed with certainty at the beginning of quarter $t + 1$. The joint density distribution of signals across firms is given by $g(\delta_A, \delta_B)$. Thus, in our model, mutual funds are informed traders, whose fully rational trading decisions can be based on private

$^6$If $\delta_{j,t+1}$ is not proportional to firm value, then pricing is sensitive to a firm’s choice of shares outstanding.
information, in addition to the liquidity needs arising from net cash inflows and the strategic consideration of the price impacts of trades on returns. One role of private information in our model is to show that incentives to pump existing holdings can be so strong that the fund manager actually trades in the opposite direction of his private information.

At the end of quarter $t$, the fund receives net cash inflows of $f(r_t)$, where $r_t$ was the fund’s portfolio return over quarter $t$. The only structure that we impose is that $f(r_t)$ is increasing in $r_t$—greater past fund performance raises net cash inflows from investors. Presumably, the performance-flow relation reflects investor beliefs that fund managers differ in their abilities to identify good investments (Berk and Green, 2004). Here, we do not distinguish fund managers by ability, because such differences can also lead to persistence in performance.

Consistent with the findings of Chan and Lakonishok (1995) and Keim and Madhavan (1997), we assume that a fund manager’s stock purchases have short-term price impacts: the more shares that a fund manager purchases, the greater is the short-term price impact. Specifically, a purchase of $I_{jt} > 0$ shares in stock $j$ costs $P_{jt} + \Delta P_{jt}(P_{jt}, I_{jt}) > P_{jt}$ per share. We summarize the properties of the price impact of $I_{jt}$, $\Delta P_{jt}(P_{jt}, I_{jt})$, later in Assumptions A1-A4. We are agnostic as to the fundamentals driving the price impact of $I_{jt}$ — it could be the fact that institutional trades contain information, or that market makers must be compensated for having to re-adjust their portfolios, and it takes time to do so. What is key is only that such institutional trading activity has a significant and persistent impact on price. It is important to note that our pricing relation does not preclude the possibility that market makers adjust their pricing to reflect the end-of-quarter incentives of funds; or that larger orders impose greater inventory costs on market makers.8

In each quarter $t$, the timing of events is as follows:

1. Firms announce period earnings and market participants revise expectations about the discounted present value of a firm’s cash flows so that stock prices equal $P_{A,t}$ and $P_{B,t}$.

2. The fund manager learns $\delta_{A,t+1}$ and $\delta_{B,t+1}$ and invests available cash to maximize the

7Bhattacharyya and Nanda (2007) augment a standard single asset static Kyle (1985) model by having an informed agent’s objective be the weighted average of (i) a short-run return based on the price generated by his trade on his holdings of the asset and (ii) the final return on the stock, when market maker’s have a signal about the agent’s pre-trade holdings of the asset. They derive a linear pricing rule that is less sensitive to order flow than the standard model.

8Hendershott and Seasholes (2007) document that, on average, prices rise when larger orders substantially reduce market maker inventories, before falling subsequently in the next week when market makers re-establish their inventories.
fund’s expected period return. Available cash consists of net cash inflows \( f(r_{t-1}) \) plus the present value of last period’s cash position \( M_{t-1} \). The fund manager can sell shares, but cannot sell shares short nor borrow to finance stock investments.

3. The fund manager’s private information about next period’s signal, \( \delta_{j,t+1} \), is revealed with (independent) probability, \( \gamma \), and is not revealed with residual probability \( (1-\gamma) \). We assume that \( \gamma \in (0,1) \), i.e., information is only sometimes revealed to the public. A smaller value of \( \gamma \) reflects less leakage of information between the purchase and the end of the period. Thus, \( \gamma \) captures either: (i) the length of time between the purchase and the end of the quarter; or (ii) the probability that private signals will be revealed (because the stock is smaller and hence is followed by fewer analysts); or (iii) a combination of both effects.

4. End-of-quarter stock prices, \( P^*_A, P^*_B \), are realized. If the fund manager’s private information was revealed,

\[
P^*_j = P_j + \delta_{j,t+1} P_j.
\]

If the private information was not revealed,

\[
P^*_j = P_j + \Delta P_j(P_j, I_j).
\]

5. End-of-quarter fund returns \( (r_t) \) are realized. The fund receives net cash inflows of \( f(r_t) \).

Figure 1 provides a sketch of the timing of events.

We next set out the sole structure imposed on the short-term price impact of share trades:

**A1.** If \( I_j = 0 \), there is no price impact: \( \Delta P_j(P_j, 0) = 0 \).

**A2.** The price impact of larger orders is greater: \( \partial \Delta P_j(P_j, I_j) / \partial I_j > 0 \).

**A3.** The price schedules are symmetric: \( \Delta P_A(\cdot) = \Delta P_B(\cdot) \).

**A4.** The price schedule is concave, but not too concave:

\[
\frac{\partial^2 \Delta P_j(P_j, I_j)}{\partial I_j^2} \leq 0; \quad 2 \frac{\partial \Delta P_j(P_j, I_j)}{\partial I_j} + I_j \frac{\partial^2 \Delta P_j(P_j, I_j)}{\partial I_j^2} > 0.
\]

Assumption **A1** is largely a normalization. Assumption **A2** captures the empirical regularity that larger orders have greater price impacts, and is an equilibrium property of economic models in which individual orders are informationally large (Kyle (1985), etc.), or are large
from the perspective of market maker inventories. Assumption A3 allows us to abstract away from how different price schedules affect a fund manager’s investment decisions, and is consistent with the assets having the same stochastic properties, and market makers not knowing the holdings of individual fund managers. Later, we allow price schedules to differ across stocks. Assumption A4 is a technical assumption whose role is to ensure that the fund manager’s optimal trading strategy is characterized by first-order conditions.

Some of our results will be most transparent when we assume A5. The price impact of an order is proportional to the value of the order:

\[ \Delta P(P_{jt}, I_{jt}) = k(P_{jt}I_{jt})P_{jt} \quad k > 0 \]  

(2)

Assumption A5 is implied if pricing is neutral with respect to a firm’s choice of shares outstanding.

Fund Manager’s Problem: At the beginning of quarter t, the fund manager invests to maximize the expected quarter portfolio return:

\[
\max_{M_t, I_{jt}, I_{rt}} E[r_t] = \frac{M_t + \sum_{j} S_{j,t+1}E[P^*_{jt}]}{M_{t-1} + f(r_{t-1}) + \sum_{j} P^*_{jt}S_{jt}} - 1
\]  

(3)

subject to

\[
M_t \geq 0, \\
I_{jt} \geq -S_{jt}, \quad j = A, B \\
\sum_{j} [P_{jt} + \Delta P_{jt}(P_{jt}, I_{jt})]I_{jt} + M_t \leq f(r_{t-1}) + M_{t-1},
\]
where $S_{j,t+1} = S_{jt} + I_{jt}$ and $E[P^*_jt] = P_{jt} + (1 - \gamma)\Delta P_{jt}(P_{jt}, I_{jt}) + \gamma\delta_{j,t+1}P_{jt}$.

We will assume that the short-selling constraint, $I_{jt} \geq -S_{jt}$, $j = A, B$ does not bind: the fund manager does not receive such a bad signal about a stock that he wants to sell more shares of the stock than he has in his portfolio.

### 3.2 Discussion of Model

**Modeling approach.** The logical construction of our model is deceivingly simple, masking the fact that it contains many ingredients that researchers have not yet been able to analyze even separately within a structural market microstructure setting. Central to our model is an informed fund manager who is budget-constrained; chooses how many shares of multiple assets to purchase or sell; makes investment decisions over time; and must incorporate the fund’s existing asset holdings into trading decisions. The existing holdings enter non-neutrally because the price impacts of trades affect the portfolio return that investors observe and condition future mutual fund investments on. Further complicating this dynamic optimization problem, a fund manager’s trades must have price impacts; trading scales cannot be trivial (i.e., zero versus single round lots); and trading rules will be highly non-linear.

There is a large market microstructure literature dating back to Kyle (1985) analyzing a single asset that generates endogenously the price impacts of orders of the form that we model, and that are found in the data.\(^9\) In Kyle and its multi-agent extensions, risk-neutral speculators receive normally-distributed signals about one asset’s terminal value, submit orders over time that are mixed in with normally-distributed exogenous “noise trade”, and a market maker sets price equal to the asset’s expected value given the net order flow. The normal assumptions imply linear conditional forecasts, and hence linear pricing, making the problem solvable. Caballé and Krishnan (1994) extend this analysis to multiple stocks in a static setting and prove that a linear equilibrium exists. Bernhardt and Taub (2008) provide rich analytical characterizations of the Caballé and Krishnan model, and they also solve for equilibrium outcomes when speculators can condition their trades on prices. To our knowledge, there are no other theoretical papers with multiple assets and price impacts.

\(^9\)Noisy rational expectations frameworks are inappropriate here, as they assume that individuals are small, so their individual trades do not generate the price impact found in the data.
Technically, the biggest challenge to handle in a model with full primitives is the budget constraint, which destroys linearity of forecasts (and hence prices) in models with normally distributed signals. Not only does our fund manager have a budget constraint, but the budget is endogenous, there are multiple risky assets, and assets-in-place critically affect optimal trades. While the budget constraint is central to the economics of our results, it is not central to the form of pricing—in any setting where an agent’s trade is “large”, the equilibrium pricing schedule will have price impacts (larger orders receive higher prices), simply because an agent with a better signal about an asset buys more, and must face an opportunity cost of buying more—in the form of a higher price.

One can also derive endogenously the qualitative consequences of introducing an investor who wants to devote resources toward assets in place within an equilibrium model. Bhattacharyya and Nanda (2007, hereafter B-N) do this by adapting Kyle (1985) to consider a single investor who cares about both the interim payoff on an existing inventory position on a single risky asset and the final payoff. The investor has a 2-period horizon. At date 1, the investor receives a private signal and trades once in a simultaneous batch auction. At date 2, the asset is liquidated at its true value. Thus, the investor’s date 1 trading decision is based on the NAV of his position at dates 1 and 2; in particular, the investor incorporates the price impact of trading on the value of his existing position at date 1. As a result, the optimal trade size is linearly related to the size of the existing position.

In the most general B-N setting, the market maker has a noisy signal of the investor’s (normally-distributed) existing position. This preserves linear forecasts and hence linear pricing. The investor’s incentives to pump up the value of his holdings reduce the sensitivity of the equilibrium price schedule to order flow, but the qualitative properties of pricing are otherwise preserved. B-N argue that the investor’s incentives to enhance short-term values can explain why closed-end funds trade at a discount to their NAVs. Their model, however, cannot shed light on any of the issues we study, as it cannot deal with multiple risky assets, the budget constraints of mutual funds, or fund managers with multiple trading opportunities.

In sum, while some of our assumptions are not grounded in primitive foundations, our framework only requires a mild reduced-form structure on pricing. This pricing structure holds empirically and allows us to get at the key links between the market structure of asset pricing, strategic fund manager behavior, and portfolio returns; thereby reconciling a host of fundamental empirical regularities in fund performance.
Empirical evidence in support of price impact function. Keim and Madhavan (1996) find that the price impact of a block trade, measured as the market-adjusted difference between the closing price on the day prior to the trade and the closing price on the day after the trade, is $-1.50\%$ for seller-initiated blocks and $1.60\%$ for buyer-initiated blocks. Chan and Lakonishok (1995) measure the price impact of a package of institutional trades and find that buy packages are associated with a principal-weighted average price change of almost one percent from the open on the package’s first day to the close on the last day. The analogous price change for sell packages is $-0.3\%$. This price impact is persistent—for example, five days after the completion of a package trade, there is a reversal in returns of only $-0.07\%$ for buy packages and $0.10\%$ for sell packages.

Frequently, institutional traders use crossing networks, such as POSIT, or use so-called dark liquidity pools to reduce the price impact of their trades. It would be interesting to investigate empirically whether mutual funds reduce their usage of these alternative trading venues near the end-of-quarters.

Time horizon of fund manager. While fund manager compensation is proportional to assets under management, a fund manager’s expected compensation may depend more on gaining a promotion based on past fund performance. Fund manager turnover is high: Baks (2007) finds that on average about 45% of managers leave or start at a fund in any given year and that the average time spent at one fund is about three years (less for small company growth funds). To capture these incentives, we suppose that each period, a fund manager invests to maximize end-of-period return. We interpret the period’s end as the date at which investors receive fund performance information, which leads them to re-allocate investments. Focusing on one-period horizons for fund managers eases the analysis, and the incentive effects highlighted remain present with longer horizons.

The one-period horizon in our model is best interpreted as a quarter, since this is the time horizon most commonly used to evaluate mutual fund performances. Our model does not intend to capture all investment decisions made within the quarter — stock purchases and sales made early in a quarter may be motivated by reasons not modeled here. The key is only that the portfolio’s end-of-quarter composition is influenced by the features that we model, and that better past performing funds tend to have more cash on hand at the quarter’s end because of higher accumulated cash inflows throughout the period. We do not claim that cash

\[^{10}\text{Hu, Hall, and Harvey (2000) find that promotions are positively related to the fund’s past performance and demotions are negatively related to past performance.}\]
inflows received earlier in the period are held in cash (or a passive equity equivalent); although we do not model it formally, fund managers presumably initially allocate that cash based on private information, etc., and then re-allocate at quarter’s end to reflect the features that we model. This modeling assumption is consistent with the empirical results of Alexander, Cici, and Gibson (2007), who show that large fund inflows and outflows influence a fund manager’s trades. It is also in the same spirit as the theoretical model of Hugonnier and Kaniel (2008), who examine portfolio choices by a mutual fund manager who faces dynamic fund flows. They show that proportional fees in combination with the performance-flow relation can cause a manager to distort a fund’s risk profile.

Definition of fully-invested funds: Most mutual funds have small cash holdings that largely reflect (i) recent cash inflows from investors that have not yet been invested; and (ii) funds held as an insurance to reduce the transaction costs associated with (stochastic) fund redemptions. Yan (2006) finds that the average cash holdings of U.S. domestic equity mutual funds is about 5.3% of assets and shows that the level of cash holdings is primarily driven by fund flow volatility. In fact, Edelen (1999) finds that about 70% of all mutual fund trades are to meet investor liquidity demands. Our model does not incorporate such investor liquidity demands, and hence it does not contain an insurance role for cash holdings. Consequently, when our model refers to fully-invested funds, we are thinking about real-world funds with cash positions no larger than optimally held as insurance against their expected redemptions.

4 Analysis

4.1 Distortion of investment towards existing holdings

The cost of the fund’s new stock purchases, \((I_{At}, I_{Bt})\), at period \(t\) are:

\[
cost_{At} = [P_{At} + \Delta P_{At}(P_{At}, I_{At})]I_{At}
\]

\[
cost_{Bt} = [P_{Bt} + \Delta P_{Bt}(P_{Bt}, I_{Bt})]I_{Bt}
\]

Differences in signals across stocks may offset the incentives to distort investment towards existing holdings; thus, to control for differences in signals, we reverse the signals when comparing investments in each stock. Specifically, consider two opposing signal patterns: \((\delta_{A,t+1} = \delta_1, \delta_{B,t+1} = \delta_2)\) and \((\delta_{A,t+1} = \delta_2, \delta_{B,t+1} = \delta_1)\). Let \(cost_{At}^{12}\) denote the cost of
optimal investment in stock $A$ under the first signal pattern, and $\text{cost}^{21}_{Bt}$ denote the cost of
optimal investment in stock $B$ under the second signal pattern; where optimal investments
are determined by the fund manager’s investment problem in (3). Our first result is:

**Proposition 1** Given assumptions A1–A5, controlling for differences in information sig-
nals across stocks, the fund manager distorts current purchases toward the stock in which he
has a larger existing position: $\text{cost}^{12}_{At} > \text{cost}^{21}_{Bt}$ when $P_{At}S_{At} > P_{Bt}S_{Bt}$.

The proof follows directly from the first order conditions of the fund manager’s problem.
Note that Proposition 1 does not rule out the possibility that the fund manager’s private
information signal for stock $B$ is so large that it dominates the portfolio pumping incentives
to purchase stock $A$. We can provide a stronger statement about investment in expectation
if the distribution of signals is such that the manager is just as likely to get a relatively
strong buy/sell signal for stock $A$ as for stock $B$. Formally, we say that the joint density
distribution of information signals is symmetric when:

$$g(y, \delta_B) = g(\delta_A = y, \cdot) \quad \forall y. \quad (6)$$

Adding this symmetry condition yields

**Corollary 1** If the joint density distribution over the fund manager’s private information
signals is symmetric, then in expectation (over possible signals) the manager invests more
money on the stock in which he has a larger existing position.

Corollary 1 follows immediately from integrating over signals pairwise.

Proposition 1 presents conditions under which the fund manager distorts his investment
toward assets in which he has larger existing positions. The assumed structure on the price
impact of order flow, $\Delta P(P_{jt}, I_{jt})$, ensures that independent of initial share prices, $P_{At}$ and
$P_{Bt}$, and initial holdings, $S_{At}$ and $S_{Bt}$, the fund manager always wants to distort investment
toward assets in which he holds larger positions. The result always holds without this
structure if (i) cash inflow is so high that the fund manager holds some cash or (ii) if, instead,
the fund manager is fully invested in the market, the sufficient condition in Proposition 1
always holds as long as either $|P_{At} - P_{Bt}|$ or $|\delta_1 - \delta_2|$ are small enough. Essentially, absent
the structure on $\Delta P(P_{jt}, I_{jt})$, we have to consider second-order effects related to the relative
value of purchases, and these depend on differences in ex-ante share prices and differences
in signals.
4.2 The likelihood of information revelation

We now reflect on how the likelihood that the fund manager’s private information is revealed affects his incentives to distort the fund’s purchases towards existing holdings. Our next result maintains the same structure on the price impact of order flow, and adds the condition

\[
P_A S_A > \frac{1 + \delta_A}{1 + \delta_B}
\]  

(7)

This condition ensures that the difference in existing holdings is large relative to the difference in information signals.

**Lemma 1** There exists an \( \epsilon > 0 \) such that Condition (7) holds if \( \delta_{A,t+1} \leq \delta_{B,t+1} + \epsilon \).

We now derive the implications of Condition (7):

**Proposition 2** Given assumptions A1-A5, new investment in the fund’s largest existing holding (stock A) is a decreasing function of the probability \( \gamma \) that information about asset-value innovations is revealed before the end of the period if and only if Condition (7) holds.

The proof follows from exploring how investment must change in response to a change in \( \gamma \) in order to maintain the first order condition of the manager’s investment problem.

The intuition is as follows. When Condition (7) holds, the manager’s private information advantage for buying stock A relative to buying stock B is small (or negative). In this case, the main incentive for the manager to distort investments towards stock A is a portfolio pumping motive. An increase in the likelihood of information revelation decreases the value of portfolio pumping, since with a higher likelihood the end-of-quarter price will reflect fundamental information, rather than the pumping activity. In contrast, when Condition (7) does not hold, the manager has such a large informational advantage for trading stock A that his primary incentive is to trade aggressively on the basis of this information. In this case, the more likely this information is to be revealed prior to the end of the quarter (and thereby incorporated into this period’s returns), the greater is the value of purchasing stock A, so that investment in stock A is an increasing function of \( \gamma \).

Adding symmetry, i.e. \( g(\cdot, \delta_B = y) = g(\delta_A = y, \cdot) \), we can extend Proposition 2 to consider expected investment, unconditional on the realization of the signal:

**Corollary 2** If the joint density distribution over the fund manager’s private information signals is symmetric and the dispersion of signals is sufficiently small, then, unconditionally,
expected investment in the stock with the larger existing position declines with the probability that the fund manager’s private information is revealed.

A smaller $\gamma$ represents a lower probability of information leakage. This lower probability can reflect both a smaller period of time between the purchase of stocks and a period’s end for the information to leak out, and it can reflect the observation that smaller stocks are less actively followed, so that there are fewer sources to disseminate the information. Hence, the proposition suggests that fund managers distort investment more toward existing holdings in smaller stocks or later in an evaluation period, where price impacts of trades are more likely to persist.

4.3 The importance of the cash constraint

We have shown that a fund manager distorts investments towards existing stock holdings. Despite this distortion, short-run performance may not be persistent. In particular, we now show that if the fund manager is not fully invested in the market, then the model cannot generate short-run persistence in fund performance.

Proposition 3 Suppose $P_{j,t-1}^* = P_{j,t}$. If the fund manager optimally does not use all available cash to invest in stocks (and thereby holds cash at the end of the period) and expected mutual fund returns exceed the risk-free rate, then the mutual fund’s period returns are a declining function of the cash available to be invested at the beginning of the period.

The proof follows from recognizing that when the manager optimally holds a cash position, the marginal value of relaxing the fund’s budget constraint (the Lagrange multiplier) is equal to the return on cash.

The expected mutual fund return exceeds the risk-free rate either if the fund manager has non-negative cash available to invest ($f \geq 0$) and stocks do not receive large negative signals ($\delta_j > \bar{\delta}_j$), or if negative cash inflows are offset by sufficiently large positive signals. This is because the fund manager can always earn the risk-free return by investing new fund inflows in cash. In our risk-neutral setting, it is appropriate to compare mutual fund returns with the risk-free rate. More generally, the rational behavior of risk averse investors implies that expected mutual fund returns must exceed a comparable risk-adjusted benchmark. This fact taken in conjunction with Proposition 3 implies that if fund managers are not fully invested
in stocks, then the model is \textbf{inconsistent} with the empirical regularity that returns exhibit short-run persistence. Indeed, in Proposition 3 we assume that \( P_{j,t-1}^* = P_{jt} \), so that returns are calculated under the assumption that the end-of-period \( t-1 \) share price reflected the value of the firm at that moment, i.e., there was no past investment distortion. If \( f(r_{t-1}) \) was higher due to past investment distortion, this would lead to even more negatively-correlated short-run returns.

\subsection*{4.4 Short-run persistence and long-run reversals}

In our model, persistence is best interpreted in terms of the persistence of relative (ranked) performance: there is short-run persistence when a highly ranked performer at \( t-1 \) tends to be highly ranked in \( t \); and lower ranked performers in \( t-1 \) tend to be lower ranked in \( t \). Because we do not model why a given fund’s return is initially higher (perhaps better past private information), we cannot compare the absolute levels of returns from one period to the next, but we can determine when the factors that we model lead a fund that does relatively better at \( t-1 \) to do better at \( t \), but eventually to do worse over sufficiently longer horizons.

In the context of our model, long-run returns are just the multi-period return realized by a sequence of decisions made by a fund manager (or multiple fund managers). Because short-run returns eventually fall for \( f(r_{t-1}) \) sufficiently large, ‘Ponzi-schemes’ cannot be supported in the long run.

Our model does not introduce persistence of fund manager ability. Here, we focus on the incentives to portfolio pump, and show that this portion of the story generates long-run mean reversion. A reversal in performance means that worst past performance implies eventual better future performance on a ranking scale. A fund over-performs by more when its rank is higher over the evaluation period that we calculate returns; and a fund under-performs by more when its return rank is lower.

We first show that the model generates short-run persistence in performance if managers are fully invested in the market. In particular, Proposition 4 proves that if a fund manager is fully invested, then short-run returns first rise with cash inflows, \( f(r_{t-1}) \).

\textbf{Proposition 4} Suppose \( P_{j,t-1}^* = P_{jt} \). Suppose that the fund manager has the same private information about each stock \( (\delta_{A,t+1} = \delta_{B,t+1} = \delta) \) and that he is fully invested in stocks.
(i.e., does not hold a cash position). Then there exists an \( f^* \), such that for \( f < f^* \), a higher return last period (and hence a higher \( f \)) implies a higher return this period; and for \( f \geq f^* \), further increases in returns in the previous period imply a lower current return.

In other words, short-run returns first rise with new funds under management (\( \partial r_t / \partial f > 0 \)) for cash inflows sufficiently small, but are a declining function of new funds for cash inflows sufficiently large. The finding for \( f < f^* \) is the relevant one, as it implies performance persistence (past higher return implies current higher return, i.e., better past performers continue to perform better).

The persistence documented in Proposition 4 is driven precisely by the flow-related buying that pushes up stock prices that Wermers (2003) documents. Observe that \( \partial r_t / \partial f > 0 \) when \( f \) is small even if a fund manager has no private information, so that \( \delta = 0 \). The intuition for Proposition 4 is sharpest when \( S_{At} \sim S_{Bt} \) and \( P_{At} \sim P_{Bt} \), in which case \( I_{At} \sim I_{Bt} \sim 0 \). Then \( I_{jt} \) shares are purchased at a premium of \( \Delta P(P_{jt}, I_{jt}) \), so the ‘cost’ \( \Delta P(P_{jt}, I_{jt})I_{jt} \) is only of second order. In contrast, the price impact of the share purchase on returns has a first order positive impact of order \( (1 - \gamma) \Delta P(P_{jt}, I_{jt})S_{jt} \).

The rest of the proof shows that the result extends when a fund manager makes non-trivial offsetting investments in the two stocks.\(^{12}\)

At date \( t + 1 \), \( \delta_{A,t+1} \) and \( \delta_{B,t+1} \) are revealed and incorporated into prices, so that, ceteris paribus, greater investment distortions at date \( t \) reduce returns in \( t + 1 \) by more. Possibly offsetting this decline is the fact that the period \( t \) investment distortion induced greater cash inflows, \( f(r_t) \), which, in turn, can facilitate another round of investment distortion. To derive the impact of past performance for long-run returns, one just cumulates short-run returns over time. While high cash inflows in period \( t \) of \( f(r_{t-1}) \) due to better performance in \( t - 1 \), lead to higher returns in period \( t \), for short-run returns in the next period \( t + 1 \) not to fall, the second round of investment distortion must dominate the impact due to the revelation of \( \delta_{A,t+1} \) and \( \delta_{B,t+1} \). For longer-run returns to continue to rise, it must be that the higher short-run return from the immediate distortion of investment more than offsets lower returns due to realizing past distortions. But short-run returns must eventually decline with

\(^{11}\)One can re-interpret Proposition 3 in the context of Proposition 5 as saying that if it is optimal for a fund not to be fully invested, then cash inflows are too high relative to private information, so that short-run returns are a decreasing function of \( f(r_{t-1}) \).

\(^{12}\)As in Proposition 3 we calculate returns using \( P_{jt,t-1}^* = P_{jt} \), i.e., higher values of \( f(r_{t-1}) \) primarily reflect factors other than greater past investment distortion.
cash inflow, so once cash inflow is sufficiently high, this cannot occur: eventually, realizing lower returns from greater past distortions must lead to lower long-run returns.

We now consider a fund manager who has no private information, so that greater incremental investments in existing positions represent more costly distortions that must eventually be realized in the form of lower returns. It follows that greater cash inflows in period $t$ must lead to lower long-run returns when cumulated over the period $[t, t+\tau]$, for $\tau$ sufficiently large. Combining this observation with Proposition 4, yields the following proposition, documenting that the model reconciles both the short-run persistence in fund performance and the long-run reversal in fund performance documented by Zheng (1999) and Wermers (2003).

**Proposition 5** Even if a fund manager has no private information, for $f(r_{t-1})$ sufficiently small, short-run returns are a rising function of $f(r_{t-1})$, but long-run returns over $[t, t+\tau]$ are a declining function of $f(r_{t-1})$, for $\tau$ sufficiently large.

Proposition 5 indicates that a better performer in $t-1$ will be higher ranked in $t$ (on the basis of returns), but lower ranked (on the basis of returns) by some date $\tau$.

Because the fund manager has no private information, from a long-run perspective, the optimal stock investment is zero. For $\tau$ sufficiently high, long-run returns are a strictly declining function of $f(r_{t-1})$, because the price premium paid for each share rises with the investment. We show later that $\tau$ need not be very large for long-run returns to fall with $f(r_{t-1})$.

These results hold even if past returns were so bad that redemptions lead to a net outflow of money from the fund. Then the mutual fund manager must disinvest. To minimize the adverse return consequences the fund manager tends to sell stocks in which he has smaller positions (i.e., the negative price impact from selling impacts a smaller proportion of his portfolio). Further, in the environment characterized by Proposition 5, short-run returns are lower (and negative) for fund managers with greater redemptions, but there will be a long-run reversal in performance. The intuition is analogous to that described before: in this situation, a fund manager sells near the end of the quarter, but then begins the next quarter with a positive return (as prices revert back to fundamental value); eventually, as more quarters pass, this positive return is large enough to offset the negative impact of fund outflows; thereby allowing the fund to become a “winner”. In general, the loser-to-winner effect will be smaller than the reverse effect, since the stocks being sold constitute a smaller proportion of the fund’s holdings.
Because cash inflows are more sensitive to fund performance for newer funds (Chevalier and Ellison, 1997), the model predicts that the persistence in short-run returns should decline as funds mature, and, in turn, there should be a smaller long-run reversal in performance for mature funds, as Zheng (1999) documents. Over time, funds with high short-run returns should under-perform the market in the long-run, as Zheng also finds.

Are mutual fund investors stupid? Our theoretical analysis and Zheng’s empirical work suggest that it is advisable to invest in a fund that just performed well for the “first” time. Our results indicate that if investors re-assess their mutual fund holding on a quarterly basis, then the optimal time to exit a fund is related to both the previous quarter’s performance, and the length of time the fund has been performing well; the more time passes, the higher the previous quarter’s performance must be to merit continuing to hold the fund.

5 Examples

Using a series of simple examples, we now illustrate how the return-cash flow relationship characterized above emerges in the short- and long-run.

5.1 Identical Stock Holdings

Consider an economy in which the fund manager has no private information, \( \delta_{A,t+1} = \delta_{B,t+1} = 0 \), and identical holdings, \( S_{At} = S_{Bt} = S > 0 \). The two identical stocks share an initial common price of \( P_{At} = P_{Bt} = 1 \) and trades have a linear price impact, \( \Delta P(1, I_{jt}) = aI_{jt}, a > 0 \).

It is straightforward to verify that there is a critical value \( \bar{f} \) such that the fund manager is fully invested if and only if cash inflows, \( f \), are less than \( \bar{f} \). Further for \( f < \bar{f} \), the fund manager optimally divides his purchases equally between stocks, purchasing \( I(f) = \frac{\sqrt{1 + 2a f}}{2a} \) shares of each stock to generate an expected mutual fund return of

\[
E[r(f)] = \frac{2(I(f) + S)(1 + (1 - \gamma)aI(f)) - (2S + f)}{2S + f}.
\]

Initially, returns from investing are increasing in \( f \),

\[
\frac{\partial}{\partial f} [E[r(f = 0)]] = \frac{a(1 - \gamma)}{2} > 0,
\]
but the second derivative with respect to \( f \) is negative,
\[
\frac{\partial^2}{\partial f^2} \left[ E[r(f = 0)] \right] = \frac{a[-(1 - \gamma)S - 1]}{2S} < 0.
\]

Figure 2 illustrates the fund flow-return relationship. The short-run portfolio return first rises and then falls with cash flow into the fund. But note that the cash inflow that maximizes short-run returns, \( \bar{f} \), is quite high, about 40% of the portfolio value. That is, although Proposition 4 only proves that short-run returns are initially increasing in cash inflow, the figure reveals that short-run returns can continue to rise even if \( f \) is substantial. Hence, there is short-run persistence in portfolio performance, as better immediate past performance draws more cash inflow, which gives rise to higher current returns. To the extent that fund managers differ in ability or there are momentum effects\(^{13}\) the short-run persistence is further magnified.\(^{14}\)

Within this symmetric framework, a comparative statics analysis yields sharp predictions:

- Short-run returns fall with the information arrival rate, \( \gamma \). That is, short-run returns are higher if information is less likely to leak out by the end of the period:
  \[
  \frac{\partial E[r(f)]}{\partial \gamma} = \frac{-2aI(f)I(f) + S}{2S + f} < 0.
  \]

- Greater cash inflows raise short-run returns by more if information is less likely to leak out:
  \[
  \frac{\partial^2 E[r(f)]}{\partial \gamma \partial f} = - \left( 4aI(\frac{\partial I(f)}{\partial f}) + 2aS \frac{\partial I(f)}{\partial f} \right) (2S + f)^{-1} + 2aI(f) (I(f) + S) (2S + f)^{-2}
  \]
  \[
  = -2a(2S + f)^{-2} \left[ I \left( 2(2S + f) \frac{\partial I(f)}{\partial f} - I(f) \right) + S \left( (2S + f) \frac{\partial I(f)}{\partial f} - I(f) \right) \right] < 0,
  \]
  because \( \frac{\partial I(f)}{\partial f} f > I(f) \) and \( S \geq 0 \).

\(^{13}\)Carhart (1997), Sapp and Tiwari (2003), and Wermers (2003) find that winning funds hold winning stocks, while the worst-performing funds tend to hold the largest positions in the worst-performing stocks. Our model predicts that funds are particularly reluctant to sell their largest positions (as Wermers (2003) finds), so that any momentum effects would magnify the persistence in short-run performance, and delay the long-term performance reversal in our model. Still, Wermers finds that such momentum effects only explain a small portion of the persistence (price impact from fund purchases explain more).

\(^{14}\)We can extend our model to consider large fund complexes in which a fund manager of a large fund within the fund family may have the incentive to buy shares of stocks in which a smaller fund within the fund family has a large stake. The cost of the newly-purchased shares has a negligible effect on the returns of the large fund, but could have a large positive effect on the small fund’s return. Given the documented return-flow relationship, the net result would be higher net cash inflows for the fund family. If this practice is widespread, we would expect to observe short-term persistence and long-term reversals in the aggregate returns of a family of funds.
These two predictions can reconcile the finding of Zheng (1999), and others, that niche mutual funds (which invest in stocks where there are fewer sources of information) have stronger short-run persistence and greater long-run reversals in performance.

5.2 Differences in Price Schedules

We now explore how outcomes are affected if the price impact of trades differs across stocks. We suppose that $\Delta P_{At}(I_{At}) = aI_{At}$ and $\Delta P_{Bt}(I_{Bt}) = bI_{Bt}$, where $a \neq b$, but that the stocks are otherwise identical, $P_{At} = P_{Bt} = 1$ and $\delta_{At} = \delta_{Bt} = \delta$. Figure 3 numerically characterizes how $I_{At}$ depends on $a$ when the fund manager is fully invested. We consider two cases: (i) balanced holdings ($S_{At} = S_{Bt} = S$); and (ii) unbalanced holdings ($S_{At} > S_{Bt}$). The figure highlights that as long as cash inflow, $f$, is sufficiently small, investment in stock $A$ rises with $a$; but for higher values of $f$, investment falls with $a$. Further, these investment consequences are magnified when $S_{At} > S_{Bt}$. Phrased differently, if and only if cash inflows are not too high, the gain from strategically manipulating returns by distorting investments toward greater existing positions rises with the price impact of stock $A$ order flow, $a$. However, if cash inflows are too high, increases in $a$ reduce investment in stock $A$ since the strategic manipulation gains are dominated by the increasing marginal cost of additional share purchases.

Next, we use our framework to predict differences in the investment behavior of specialized funds. To do so, we equate the price impact of order flow for the two stocks ($a = b$), and then consider how investment, stock prices, and fund returns change as we vary this parameter. Figure 4 illustrates outcomes when $S_{At} = 30$ and $S_{Bt} = 10$. Since the only distinction between the two stocks is the existing holdings, without the short-term painting the tape motive, the long-term optimal investment would be to purchase the same amount of both stocks. Panel A shows that as stock price sensitivity increases, the amount of investment distortion (as measured by the difference between $I_{At}$ and $I_{Bt}$) falls. At the same time, panel B shows that the price movement in the more sensitive (less liquid) stocks is larger. Thus, our model predicts that a fund with holdings in small, illiquid stocks will have greater return distortions, but less investment distortion, than an otherwise comparable fund with holdings in larger, more liquid stocks. As well, the apparent clairvoyant stock picking effect should be reduced for larger, more liquid stocks since these stocks will have less price impact from distortionary trades.
5.3 Returns over multiple periods, different initial stock holdings

The numerical findings illustrated in Figure 2 do not imply that higher returns will persist for some time. This is because a fund must realize the negative return consequences of its immediate past investment distortions each period, as share prices incorporate the true value of past signals. Figure 5 considers longer-period returns when cash inflows in each period \( t \) are given by \( f(r_{t-1}) = [0.2 + 4.2(r_{t-1} - \delta)] \times (P_{A,t-1}S_{A,t-1} + P_{B,t-1}S_{B,t-1} + M_{t-1}) \), where \( \delta = \delta_A = \delta_B \forall t \). This flow-performance relation is a linear approximation of the relation estimated by Chevalier and Ellison (1997) for 2-year old mutual funds, and subsequently modeled by Berk and Green (2004). We illustrate period 1 fund flows ranging from \(-10\%\) to \(40\%\) of existing holdings, which is roughly consistent with the typical range of annual fund flows observed by Chevalier and Ellison (1997) for 2-year old mutual funds. The figure highlights that better-performing funds in period 0, which consequently generate high cash inflows at the beginning of period 1, also do better in period 1: there is very strong short-run persistence in performance. However, in period 2, past better performers under-perform by amounts that substantially offset their superior period 1 performance, and were we to continue for another period, there would be a long-run reversal in performance. Obviously, if we incorporated other persistent factors, such as fund manager ability, then long-run reversals would take longer and be less extreme.

According to our flow-performance relation, the initial range of cash inflows \((-10\%\) to \(40\%\)) implies a difference in period 0 returns of 0.3\% to 30.0\%. This dispersion of returns is much larger than the dispersion of observed period 1 returns obtained using our model: 5.7\% to 12.5\%. This suggests that while painting the tape drives short-run persistence and long-run reversal patterns, other exogenous factors such as idiosyncratic stochastic shocks drive much of the dispersion in observed returns. In particular, this confirms that taking \( r_0 \) and hence initial cash flows as exogenous as we do in our propositions is unimportant as our qualitative findings would extend.\(^{15}\)

There is significant anecdotal and empirical evidence that suggests funds conduct much of their trades near the end of a quarter. In addition to the trading behavior that we model, there is evidence of “window dressing” (selling the fund’s lackluster holdings before

\(^{15}\)In unreported results, we confirmed that our qualitative findings are not affected by starting values by iterating our model over an arbitrary number of periods and verifying that the same short-run persistence and long-run reversal patterns exist in later periods.
they are revealed to investors in end-of-quarter reports). If selling poorly performing stock holdings near the end of quarters is prevalent, then fund managers will have more cash on hand than normal to make new investments near the end of quarters. Our model predicts that fund manager’s will tend to use this cash to buy shares that magnify the value of the fund’s existing holdings (i.e., stocks in which the fund has many shares or stocks that are thinly traded). This behavior has very different implications for mutual fund return patterns than the window dressing motive of buying winners (in which a fund may or may not have a substantial existing position). Window dressing alone cannot generate the short-run persistence and long-run reversal patterns in the data.\textsuperscript{16}

6 Conclusion

This paper provides a theoretical framework for understanding the incentives for, and the implications of, mutual fund managers distorting end-of-quarter stock purchases toward stocks in which the fund has larger existing positions. We show how this common practice, sometimes known as painting the tape, depends on the short-term price impact of trades and the relation between past fund performance and future fund cash inflows.

We derive when and how such trading leads to the empirically observed short-run persistence and long-run reversal in fund performance. Our model also explains why mutual funds tend to be relatively undiversified and exhibit persistent stock selection. Our theoretical predictions are consistent with the empirical findings of Carhart et al. (2002) that trading-induced equity price inflation on the last day of a quarter gives rise to abnormal fund returns on those days and that the end of year performance effect is more pronounced for better historical performers.

Our unified framework for analyzing mutual fund investment distortion provides several directions for future empirical research. For instance, fund managers will invest more heavily on the basis of private information earlier in an evaluation period and this should be reflected

\textsuperscript{16}So, too, Carhart et al. (2002) find that mutual fund net asset value inflation in excess of the S&P 500 index on the last day of a quarter is at least 0.2\% to 0.4\% for funds in lesser deciles. Carhart et al. argue that this is caused by a “leaning-for-the-tape” motive, in which fund managers near the top of the return rankings have the most to gain from performance rank improvements (as consistent with the empirical findings of Ippolito (1992) and Sirri and Tufano (1998)) and thus have the most to gain from end-of-quarter distortionary trades. But this leaning-for-the-tape motive alone cannot generate short-term persistence in mutual fund returns because there is no link between past and current period fund performance.\textsuperscript{16}
in the long-run return characteristics of these purchases. As well, the short-run persistence, long-run reversal pattern should be exhibited only by funds that are effectively fully-invested in the market (excluding cash held for insurance against unexpected redemptions). And, funds with a more sensitive performance-flow relation should exhibit greater investment distortion. Finally, we argue that existing measures of mutual fund managers’ true stock picking ability have been understated because of their failure to recognize the additional costs of the strategic behavior modeled in our paper.
A Proofs

We note that maximizing portfolio returns also maximizes the portfolio end-of-period value. Thus, the Lagrangian corresponding to the manager’s problem is

\[ \mathcal{L}(M_t, I_{At}, I_{Bt}) = M_t + \sum_j (I_{jt} + S_{jt})E \left[ P^*_{jt} \right] + \lambda_M M_t + \lambda_A [S_{At} + I_{At}] + \lambda_B [S_{Bt} + I_{Bt}] + \lambda_{bud} [f(r_{t-1}) + M_{t-1} - [P_{At} + \Delta P_{At}(P_{At}, I_{At})]I_{At} - [P_{Bt} + \Delta P_{Bt}(P_{Bt}, I_{Bt})]I_{Bt} - M_t]. \]

Given that \( \lambda_M = \lambda_A = \lambda_B = 0 \), the associated first order condition characterizing the optimal investment in stock \( j \in \{A, B\} \) is:

\[ \frac{\partial \mathcal{L}}{\partial I_{jt}} = P_{jt} + (1 - \gamma)\Delta P_{jt}(P_{jt}, I_{jt}) + \gamma \delta P_{jt} + (1 - \gamma)I_{jt} \frac{\partial \Delta P_{jt}}{\partial I_{jt}} + (1 - \gamma)S_{jt} \frac{\partial \Delta P_{jt}}{\partial I_{jt}} - \lambda_{bud} \left[ P_{jt} + \Delta P(P_{jt}, I_{jt}) + I_{jt} \frac{\partial \Delta P_{jt}}{\partial I_{jt}} \right] = 0. \]

**Proof of Proposition 1:** From the first-order conditions:

\[
\lambda_{bud} = \frac{P_{At} + (1 - \gamma)\Delta P(P_{At}, I_{At}) + \gamma \delta P_{At} + (1 - \gamma)(I_{At} + S_{At})\frac{\partial \Delta P}{\partial I_{At}}}{P_{At} + \Delta P(P_{At}, I_{At}) + I_{At} \frac{\partial \Delta P}{\partial I_{At}}} = \frac{P_{Bt} + (1 - \gamma)\Delta P(P_{Bt}, I_{Bt}) + \gamma \delta P_{Bt} + (1 - \gamma)(I_{Bt} + S_{Bt})\frac{\partial \Delta P}{\partial I_{Bt}}}{P_{Bt} + \Delta P(P_{Bt}, I_{Bt}) + I_{Bt} \frac{\partial \Delta P}{\partial I_{Bt}}},
\]

Substituting for \( \Delta P(P_j, I_j) = k(P_jI_j)P_j \),

\[
(1 - \gamma) + \frac{\gamma + \gamma \delta_1 + P_{At}S_{At}(1 - \gamma)k}{1 + 2kP_{At}I_{At}} = (1 - \gamma) + \frac{\gamma + \gamma \delta_2 + P_{Bt}S_{Bt}(1 - \gamma)k}{1 + 2kP_{Bt}I_{Bt}}
\]

Then, reversing \( P_{At}S_{At} \) and \( P_{Bt}S_{Bt} \), holding \( P_{At}I_{At} \) and \( P_{Bt}I_{Bt} \) fixed, it is clear that:

\[
\frac{\gamma + \gamma \delta_1 + P_{Bt}S_{Bt}(1 - \gamma)k}{1 + 2kP_{At}I_{At}} < \frac{\gamma + \gamma \delta_2 + P_{At}S_{At}(1 - \gamma)k}{1 + 2kP_{Bt}I_{Bt}}
\]

To restore equality, it follows that the optimal values of \( I_{At} \) and \( I_{Bt} \) must satisfy \( P_{At}I_{At}(\delta_1, \delta_2) > P_{Bt}I_{Bt}(\delta_2, \delta_1) \). Then it follows that

\[
(P_{At} + \Delta P(P_{At}, I_{At}(\delta_1, \delta_2)))I_{At}(\delta_1, \delta_2) > (P_{Bt} + \Delta P(P_{Bt}, I_{Bt}(\delta_2, \delta_1)))I_{Bt}(\delta_2, \delta_1).
\]

**Proof of Proposition 2:** Re-arranging (10) gives:

\[
\frac{\gamma(1 + \delta_A - kP_{At}S_{At}) + kP_{At}S_{At}}{\gamma(1 + \delta_B - kP_{Bt}S_{Bt}) + kP_{Bt}S_{Bt}} = \frac{1 + 2kP_{At}I_{At}}{1 + 2kP_{Bt}I_{Bt}} \]

(11)
Taking the derivative of the left hand side of (11) with respect to $\gamma$ gives:

$$
\frac{k(P_B S_B - P_A S_A) - \delta_A P_B S_B - \delta_B P_A S_A k}{[\gamma(1 + \delta_B - kP_B S_B) + kP_B S_B]^2}
$$

Since $P_A S_A > P_B S_B$, this derivative is negative if $\delta_A \leq \delta_B$ or more generally iff condition (7) holds. Since the right hand side of (11) is increasing in $P_{At} I_{At}$, our desired result follows directly. 

**Proof of Proposition 3:** If $M_t > 0$, then $\lambda_{bad} = 1$ and the marginal dollar is invested in cash ($dM_t/df = 1; \partial I_j/df = 0$, $j = A, B$). Differentiating short-run expected returns, 

$$
E[r_t] = [M_t + E[P^*_A (I_{At} + S_{At}) + E[P^*_B (I_{Bt} + S_{Bt})]] [P_{At} S_{At} + P_{Bt} S_{Bt} + f(r_{t-1})] - 1,
$$

with respect to $f = f(r_{t-1})$ at the manager’s optimal values of $I_{At}$ and $I_{Bt}$ yields

$$
\frac{\partial E[r(I_{At}, I_{Bt})]}{\partial f} = \left[ \frac{dM_t}{df} + \frac{d}{df} \left( E[P^*_A (I_{At} + S_{At}) + E[P^*_B (I_{Bt} + S_{Bt})]\right) \right] [P_{At} S_{At} + P_{Bt} S_{Bt} + f]^{-1} 
- \left[ M_t + E[P^*_A (I_{At} + S_{At}) + E[P^*_B (I_{Bt} + S_{Bt})][P_{At} S_{At} + P_{Bt} S_{Bt} + f]^{-2} 
= [P_{At} S_{At} + P_{Bt} S_{Bt} + f]^{-1} 
- \left[ M_t + E[P^*_A (I_{At} + S_{At}) + E[P^*_B (I_{Bt} + S_{Bt})][P_{At} S_{At} + P_{Bt} S_{Bt} + f]^{-2} 
= -E[r_t][P_{At} S_{At} + P_{Bt} S_{Bt} + f]^{-1} \leq 0, \text{ if } E[r_t] \geq 0; \text{ strict if } E[r_t] > 0.

where the last equality follows from substitution. 

**Proof of Proposition 4:** From the budget constraint,

$$
f = [P_{At} + \Delta P_{At}(P_{At}, I_{At})] I_{At} + [P_{Bt} + \Delta P_{Bt}(P_{Bt}, I_{Bt})] I_{Bt}, \quad (12)
$$

$$
1 = \frac{\partial I_{At}}{\partial f} (P_{At} + \Delta P_{At}) + \frac{\partial \Delta P_{At}}{\partial I_{At}} I_{At} + \frac{\partial I_{Bt}}{\partial f} (P_{Bt} + \Delta P_{Bt}) + \frac{\partial \Delta P_{Bt}}{\partial I_{Bt}} I_{Bt}.
$$

We first prove the result that short-run returns are increasing in $f$ when $I_{At}$ and $I_{Bt}$ are small if $f$ is close to zero (as will be the case for $S_{At} \sim S_{Bt}$, and $P_{At} \sim P_{Bt}$). Then, it follows that:

$$
1 = \frac{\partial I_{At}}{\partial f} P_{At} + \frac{\partial I_{Bt}}{\partial f} P_{Bt}.
$$
If the fund manager is fully invested, then expected returns are

\[ E[r_t] = \frac{\sum_{j=\{A,B\}} \{(1 + \gamma \delta)P_{jt}(I_{jt} + S_{jt}) + (1 - \gamma)\Delta P_{jt}(P_{jt}, I_{jt})(I_{jt} + S_{jt})\}}{P_{At}S_{At} + P_{Bt}S_{Bt} + f} - 1. \] (13)

Thus,

\[ \frac{\partial E[r_t]}{\partial f}\bigg|_{f=0} = [P_{At}S_{At} + P_{Bt}S_{Bt}]^{-1} \left(1 - \gamma \right) \left[ \frac{\partial \Delta P_{At}}{\partial I_{At}} \frac{\partial I_{At}}{\partial f} S_{At} + \frac{\partial \Delta P_{Bt}}{\partial I_{Bt}} \frac{\partial I_{Bt}}{\partial f} S_{Bt} \right] + (1 + \gamma \delta) \left( \frac{\partial I_{At}}{\partial f} P_{At} + \frac{\partial I_{Bt}}{\partial f} P_{Bt} \right) \] - \left[ \frac{(1 + \gamma \delta)(P_{At}S_{At} + P_{Bt}S_{Bt})}{(P_{At}S_{At} + P_{Bt}S_{Bt})^2} \right]. \] (14)

Substituting for \( P_{Bt} \frac{\partial I_{Bt}}{\partial f} + P_{At} \frac{\partial I_{At}}{\partial f} = 1 \) yields

\[ \frac{\partial E[r_t]}{\partial f}\bigg|_{f=0} = [P_{At}S_{At} + P_{Bt}S_{Bt}]^{-1} \left(1 - \gamma \right) \left[ \frac{\partial \Delta P_{At}}{\partial I_{At}} \frac{\partial I_{At}}{\partial f} S_{At} + \frac{\partial \Delta P_{Bt}}{\partial I_{Bt}} \frac{\partial I_{Bt}}{\partial f} S_{Bt} \right] > 0 \] (15)

since \( \frac{\partial \Delta P_{At}}{\partial I_{At}} > 0 \) (by assumption) and \( \frac{\partial \Delta P_{Bt}}{\partial I_{Bt}} > 0 \) (since (10) can only hold if an increase in \( I_{At} \) coincides with an increase in \( I_{Bt} \)).

More generally to show that the derivative of returns with respect to \( f \) is positive recognize that this amounts to showing that

\[ \text{sign}\left([(P_{At}S_{At} + P_{Bt}S_{Bt}) \frac{\partial \text{num}}{\partial f} - \text{num}]\right|_{f=0} > 0, \]

where \( \text{num} \) is the numerator of expected returns. Equivalently, dividing through by \( P_{At}S_{At} + P_{Bt}S_{Bt} \), it amounts to showing that

\[ \text{sign}\left[\frac{\partial \text{num}}{\partial f} - (1 + E[r_t])\right]\bigg|_{f=0} = \text{sign}\left[\lambda_{bud} - (1 + E[r_t])\right]\bigg|_{f=0} > 0, \]

where the equality follows from substitution of equations (8) and (9). That is, \( \lambda_{bud} \) reflects the marginal return of one more dollar, and if the marginal return exceeds \( 1 + r_t \), then \( r_t \) must be increasing in \( f \). The Lagrange multiplier can be interpreted as the marginal value of one more dollar of investment in stock \( A \) or as the reduction in the marginal cost of selling one more dollar of investment in stock \( B \). We now show that for \( \delta \geq 0 \), this marginal cost grows monotonically as the fund manager sells more \( I_B \). Differentiating (9) with respect to \( I_B \) we see that it has sign

\[
(P_{Bt} + \Delta P(P_{Bt}, I_{Bt}) + I_{Bt}\Delta P'(P_{Bt}, I_{Bt})) \\
\times [(1 - \gamma)(2\Delta P'(P_{Bt}, I_{Bt}) + I_{Bt}\Delta P''(P_{Bt}, I_{Bt})) + S_{Bt}(1 - \gamma)\Delta P''(P_{Bt}, I_{Bt})] \\
- (2\Delta P'(P_{Bt}, I_{Bt}) + I_{Bt}\Delta P''(P_{Bt}, I_{Bt})) \\
\times [P_{Bt} + (1 - \gamma)\Delta P(P_{Bt}, I_{Bt}) + \gamma\delta P_{Bt} + (1 - \gamma)(I_{Bt} + S_{Bt})\Delta P'(P_{Bt}, I_{Bt})] \\
< \ (P_{Bt} + \Delta P(P_{Bt}, I_{Bt}) + I_{Bt}\Delta P'(P_{Bt}, I_{Bt}))S_{Bt}(1 - \gamma)\Delta P''(P_{Bt}, I_{Bt}) \\
\times [(2\Delta P'(P_{Bt}, I_{Bt}) + I_{Bt}\Delta P''(P_{Bt}, I_{Bt}))(\gamma\delta P_{Bt} + S_{Bt}(1 - \gamma)\Delta P'(P_{Bt}, I_{Bt}))] < 0, \]

27
where the last inequality follows from $\delta \geq 0$ and $A_4$. $\lambda_{bud}$ takes its value evaluated at $I_{Bt}^*$, so that it exceeds the derivative evaluated at $I_{Bt} = 0$, which we have proven exceeds $r_t$. For $\delta < 0$ an analogous exercise on (8) establishes monotonicity for $I_{At} > 0$.

Finally, short-run returns must eventually decline with $f$: $A_4$ implies that marginal returns from making arbitrarily large purchases of a stock must eventually become negative. This implies that once $f$ grows sufficiently large, the fund manager begins to invest in cash at the point where the marginal return on investing in stock is $i$ (which we normalized to zero), and where the short-run portfolio return exceeds $i$. It follows that for $f$ larger, short-run returns decline. ■
References


Figure 2: Fund flow-return relationship in base example. Parameters: $S_A = S_B = 20$, $a = b = 0.02$, $\gamma = 0.5$, $P_A = P_B = 1$, $\delta_{A,t+1} = \delta_{B,t+1} = \delta$. 
Figure 3: Fund manager’s investment decision and differences in price sensitivity. Parameters: $\gamma = 0.5$, $P_{At} = P_{Bt} = 1$, $\delta_{At,t+1} = \delta_{B,t+1} = 0.07$, $b = 0.02$. Balanced corresponds to $S_{At} = S_{Bt} = 20$. Unbalanced corresponds to $S_{At} = 30$ and $S_{Bt} = 10$. For both the balanced and unbalanced cases, we plot the fund manager’s investment in stock A, $I_{At}$, versus the price impact of stock A to order flow ($a$) for low cash inflows ($f = 4$) and for high cash inflows ($f = 18$).
Panel A: Investment behavior

![Graph showing net purchase of stock \((I_{At}, I_{Bt})\) vs. price sensitivity \((a = b)\)]

Panel B: Short-term price changes

![Graph showing expected price change vs. price sensitivity \((a = b)\)]

Figure 4: Investment behavior of specialized funds that focus on stocks with a given price sensitivity. Parameters: \(\gamma = 0.5\), \(P_{At} = P_{Bt} = 1\), \(\delta_{A,t+1} = \delta_{B,t+1} = 0.07\), \(f = 4\), \(S_{At} = 30\) and \(S_{Bt} = 10\). Both stocks have the same price sensitivity \((a = b)\). The fund is fully invested in stocks for the parameters considered. The expected price change for stock \(j\) is \([\gamma a(P_{jt}I_{jt}) + (1 - \gamma)\delta_{jt+1}]P_{jt}\).
Period \( t = 1, 2 \) fund return \( (r_t) \)

![Graph showing multi-period example illustrating short-term persistence and long-run reversal of performance.]

Period 0 winners are funds for which \( r_0 > i_0 \). The y-axis plots mutual fund returns \( (r_t, t = 1, 2) \) and the x-axis plots period 1 cash inflows \( f_1(r_0) \). Parameters: \( a = b = 0.02 \), \( S_{A,1} = 30 \), \( S_{B,1} = 10 \), \( P_{A,1} = P_{B,1} = 1 \), \( \gamma = 0.5 \); \( \delta_{A,t} = \delta_{B,t} = \delta = 0.07 \ \forall t \).