Window Dressing in Mutual Funds

Vikas Agarwal
Robinson College of Business, Georgia State University

Gerald D. Gay
Robinson College of Business, Georgia State University

Leng Ling
Bunting College of Business, Georgia College & State University

We provide a rationale for window dressing wherein investors respond to conflicting signals of managerial ability inferred from a fund's performance and disclosed portfolio holdings. We contend that window dressers make a risky bet on their performance during a reporting delay period, which affects investors' interpretation of the conflicting signals and hence their capital allocations. Conditional on good (bad) performance, window dressers benefit (suffer) from higher (lower) investor flows compared with non–window dressers. Window dressers also show poor past performance, possess little skill, and incur high portfolio turnover and trade costs, characteristics which in turn result in worse future performance. (*JEL* G11, G23)

An alleged agency problem in the mutual fund industry involves managers altering or distorting their portfolios in an attempt to mislead investors about their true ability by disclosing disproportionately higher (lower) holdings in stocks that have done well (poorly) over a reporting period. This practice, commonly referred to as window dressing, has the potential to adversely

---

Vikas Agarwal acknowledges research support from the Centre for Financial Research (CFR) at the University of Cologne and from Georgia State University. Leng Ling acknowledges research grant support from Georgia College & State University. We thank Ranadeb Chaudhuri, Mark Chen, Conrad Ciccotello, Gjergji Cici, K. J. Martijn Cremers, Elroy Dimson, Jesse Ellis, Wayne Ferson, Jason Greene, Zhishan Guo, Zoran Ivkovic, Marcin Kacperczyk, Jayant Kale, Aneel Keswani, Omesh Kini, Bing Liang, Reza Mahani, Ernst Maug, David Musto, Tiago Pinheiro, Chip Ryan, Thomas Schneeweis, Clemens Sialm, Vijay Singal, Tao Shu, Daniel Urban, Qinghai Wang, and Chong Xiao for their helpful comments and constructive suggestions. We benefited from the comments received at presentations at the Bank of Canada, Cass Business School, the Financial Intermediation Research Society (FIRS) 2012 Conference, the Financial Management Association 2012 Conference, the Midwest Finance Association 2013 Conference, the Southern Finance Association 2012 Conference, the University of Alabama, the University of Cambridge, the University of Georgia, the University of Mannheim, the University of Massachusetts Amherst, and Wuhan University. We acknowledge the research assistance of Sujuan Ma, Jinfei Sheng, and Haibei Zhao and thank Linlin Ma and Yuehua Tang for sharing some of the data used in the empirical analysis. We are grateful to Gang Hu, Paul Irvine, and Kumar Venkataraman for their help with the Abel Noser data and computation of trade costs. Finally, we wish to thank the editor (Laura Starks) and the referee for many valuable comments and suggestions. Supplementary data can be found on *The Review of Financial Studies* web site. Send correspondence to Leng Ling, Georgia College & State University (GCSU), Bunting College of Business, Suite 414, Milledgeville, GA 31061; telephone: (478) 445-2587. E-mail: leng.ling@gcsu.edu.

doi:10.1093/rfs/hhu045 Advance Access publication July 15, 2014

affect fund value through unnecessary portfolio churning.[1] Despite some evidence consistent with window-dressing behavior (e.g., see Lakonishok et al. 1991; Sias and Starks 1997; He, Ng, and Wang 2004; Ng and Wang 2004; Meier and Schaumburg 2004), there is limited understanding of the incentives for managers to engage in window dressing. Presumably, such incentives could be garnered from analyzing investors' reaction to such behavior. These incentives, however, present an interesting enigma unaddressed in the literature. If investors are misled by funds' window-dressing activity and thus reward such funds with higher flows, one might then ask why all fund managers do not engage in such activity. In contrast, if investors are not deceived by window dressing and punish such funds with lower flows, then why would any manager engage in it? In other words, how can we explain window-dressing behavior in the presence of rational investors? In addition to understanding this enigma is the challenge of how one can detect window dressing.

Our paper attempts to address these issues. We portray a rationale for window dressing wherein investors can receive conflicting signals regarding managerial ability, notably in the case where a fund's disclosed holdings at quarter-end do not conform to the fund's performance over the quarter. A critical feature of this rationale is that reporting requirements allow portfolio holdings to be disclosed with a delay of up to 60 days following quarter-end. We contend that a fund's subsequent performance during this delay period can affect investors' interpretation of the two conflicting signals. To illustrate, a poorly performing manager may decide to window-dress toward quarter-end and thus rebalance to disclose disproportionately higher and lower proportions of winner and loser stocks during the quarter, respectively. If the manager then performs well during the delay period, investors are less likely to attribute the signal conflict to window dressing and more likely to attribute it to improved security selection. As a result, subsequent to the delay period, managers may benefit from incrementally higher flows than what would be justified by the fund's performance. In contrast, if the performance during the delay period is bad, then investors are more likely to attribute the signal conflict to window dressing and thus cause the manager to incur the cost of incrementally lower flows. Figure 1 illustrates a timeline of the events related to the observance of the delay period performance and investor flows.

In essence, our explanation suggests that window-dressing (henceforth WD) managers are making a risky bet that will pay off if their performance during the delay period turns out to be good. Investors are more likely to believe

---

[1] In addition to the performance-based window dressing (e.g., buying winners and selling losers) that we study, the literature notes other forms of window dressing. Prior to reporting, managers may (1) decrease their holdings in high-risk securities to make their portfolios appear less risky (Musto 1997 and 1999; Morey and O'Neal 2006); (2) purchase stocks already held to drive up stock prices and thereby fund values, a practice known as "portfolio pumping," "leaning for the tape," or "marking up" (Carhart et al. 2002; Agarwal, Daniel, and Naik 2011; Hu et al. 2014); (3) invest in securities that deviate from their stated fund objectives and later sell them (Meier and Schaumburg 2004); and (4) invest in stocks covered in the media (Solomon, Soltes, and Sosyura 2014).

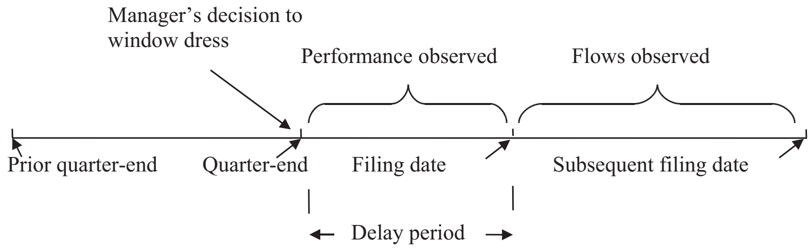

Figure 1
Timeline for the rationale of window dressing
This figure shows the timeline for the different dates and events used to explain window dressing in the presence of rational investors.

that these managers have stock selection ability if they attribute the good fund performance to the disclosed high (low) proportion of winning (losing) stocks. In this scenario, as the signals of managerial ability from both good performance over the delay period and the composition of portfolio holdings tilted toward winners reinforce each other, investors will reward such funds with higher flows. In contrast, if such managers experience continued poor performance during the delay period, then investors will suspect WD behavior and shun such funds by withdrawing or not investing capital.

We examine investor flows and find results that are consistent with our purported rationale. We find that conditional on good performance during the delay period, window dressers benefit from higher flows compared with non–window dressers. In contrast, with bad performance, window dressers incur a cost in terms of lower flows. Further, window dressers exhibit greater dispersion in flows across the two states (good and bad performance) than do non–window dressers. This supports the notion that window dressers are making a risky bet on performance during the delay period where the payoffs are in the form of investor flows.

These findings lead us to two additional research questions: (1) What types of funds or managers may engage in window dressing and make such a risky bet? and (2) Is window dressing indeed a value-destroying activity and associated with lower future performance? To address these questions, we encounter the two challenges of how to measure window dressing and how to discern window dressing from momentum and other information-based trading strategies.

Since window dressing is unobservable, an important challenge is to develop proxies that can detect it. We construct two measures. Our first measure is *Rank Gap*, which captures the discrepancy between a performance-based ranking of a fund and a ranking based on the proportions of winner and loser stocks disclosed by the fund at quarter-end. The intuition underlying the design of this measure is that, on average, a poorly performing fund should have a greater percentage of its assets invested in loser stocks and a lesser percentage invested in winner stocks than that of a well-performing fund. Thus, observing a poorly performing fund with a high percentage of disclosed holdings in winners and a low percentage in losers would suggest a greater likelihood of WD behavior. Since *Rank Gap* is based on ranking a fund's performance as well as its winner and loser proportions relative to other funds, it can be viewed as a *relative* measure.

Our second measure is *Backward Holding Return Gap* (*BHRG*), an *absolute* measure that compares the hypothetical return of a fund's reported holdings with the fund's actual return. This measure is motivated by the work of Kacperczyk, Sialm, and Zheng (2008) (henceforth KSZ), who compare a fund's actual performance with the performance of the fund's prior quarter-end portfolio, assuming it to be held throughout the current quarter. They refer to the difference between the two returns as "return gap" and attribute it to manager skill. Since we are studying WD behavior, we use instead the current quarter-end portfolio and assume that a manager held it from the beginning of the current quarter. The intuition is that a WD manager upon observing winner and loser stocks toward the quarter-end will tilt portfolio holdings toward winner stocks and away from loser stocks to give investors a false impression of stock selection ability. We thus compute *BHRG* as the difference between (a) the return imputed from the reported quarter-end portfolio (assuming that the manager held this same portfolio at the beginning of the quarter) after adjusting for trade costs and expenses; and (b) the fund's actual quarterly return.[2]

Using our measures, we investigate (a) the characteristics of funds and managers that engage in window dressing, and (b) the effects of window dressing on future performance. We find that WD behavior is associated with managers having poor performance and lacking skill, which perhaps justifies why certain managers choose to engage in the risky bet associated with window dressing. Interestingly, this finding resonates well with the literature documenting a positive association between career concerns of managers and their risk-taking behavior (see Khorana 1996; Brown, Harlow, and Starks 1996; Chevalier and Ellison 1997). We also find that WD managers engage in excessive turnover of stocks in both their WD-related and other trading. Using Abel Noser institutional transaction data, we show that window dressing results in significantly higher levels of trade costs. Further, separating window dressers' trade costs into explicit and implicit components, we find the implicit component to be particularly higher, consistent with window dressers having an urgency to trade (buy winners and sell losers) near quarter-ends.

Given that WD managers appear to be unskilled, follow non-information-based trading strategies, and incur high levels of trade costs, we conjecture that their future performance should be poor on average. Controlling for

[2] To help demonstrate the distinction between our *BHRG* measure and KSZ's return gap measure, we provide a numerical illustration in the Supplementary Appendix (available on the *RFS* Web site) that shows how *BHRG* helps identify a WD manager while the return gap measure helps identify a skilled manager.

past performance, we find that both short-term and long-term future fund performance is negatively related to window dressing.

Our findings that window dressing is associated with poor past performance as well as lower future performance show that the *BHRG* and *Rank Gap* measures are capturing window dressing rather than momentum trading. To help further make this distinction, we conduct two seasonality tests. We first examine the intraquarter variation in a fund's exposure to recent winners with the intuition that the exposure for momentum traders should be uniformly distributed due to the monthly rebalancing or updating of winners inherent in the strategy. In contrast, for window dressers the exposure should increase in the third month of the quarter due to the purchase of winners toward quarter-end. In the second test, we examine the intraquarter variation in WD behavior for December versus other fiscal quarter month-ends. There are two reasons for this test. First, the literature on tournaments and the flow-performance relation (e.g., Brown, Harlow, and Starks 1996; Chevalier and Ellison 1997; Sirri and Tufano 1998; Huang, Wei, and Yan 2007) suggests that many investors evaluate funds on a calendar-year basis, which may provide greater incentives to window-dress in December. Second, window dressers may disguise their behavior by selling losing stocks in December and thus pooling themselves with tax-loss sellers. Therefore, window dressing should be more pronounced in December while momentum trading should be more uniformly distributed over the year. The findings from these tests further corroborate that our measures are capturing window dressing and not momentum trading.

Our paper also builds on a broader literature that studies the effects of portfolio disclosure on the investment decisions of money managers (Musto 1997, 1999); the consequences of portfolio disclosure, such as free riding and front-running (Wermers 2001; Frank et al. 2004; Brown and Schwarz 2011; Verbeek and Wang 2013); the determinants of portfolio disclosure and its effect on performance and flows (Ge and Zheng 2006); the motivation behind institutions seeking confidentiality for their 13F filings (Aragon, Hertzel, and Shi 2013; Agarwal et al. 2013); and the effect of mandatory portfolio disclosure on stock liquidity and on the performance of disclosing mutual funds (Agarwal et al. Forthcoming).

1. Related Literature and Testable Hypotheses

1.1 Related literature

One strand of related literature studies the relation between the turn-of-the-year effect and window dressing by institutional investors. Haugen and Lakonishok (1988) and Ritter and Chopra (1989) argue that window dressing can potentially explain the January effect. Sias and Starks (1997), Poterba and Weisbenner (2001), and Chen and Singal (2004) attempt to disentangle tax-loss selling and WD explanations for the turn-of-the-year effect and provide evidence in support of tax-loss selling. Starks, Yong, and Zheng (2006) sharpen tests in these prior studies by studying municipal bond closed-end funds to provide further support for tax-loss selling driving the January effect.

Another strand of literature studies the trading behavior of institutions around quarter-ends to find evidence suggestive of window dressing. Lakonishok et al. (1991) examine the quarterly equity holdings of pension funds and show that they sell more losers in the fourth quarter of the year. Similarly, He, Ng, and Wang (2004) show that institutions who invest on behalf of clients sell more poorly performing stocks during the fourth quarter. Moreover, this trading behavior is more pronounced for institutions whose portfolios have underperformed. Ng and Wang (2004) find that institutions sell more extreme losing small stocks in the fourth quarter. Meier and Schaumburg (2004) propose shape tests for alternative trading patterns and find evidence consistent with window dressing in equity mutual funds.

Finally, in a more recent study, G. Hu et al. (2014) analyze the equity transactions of institutions (e.g., mutual fund families and plan sponsors) that report to Abel Noser. Although their main focus is on portfolio pumping, they also study the practice of buying winners and selling losers at calendar quarter-ends and conclude that window dressing is not a widespread phenomenon. While their finding may at first appear contrary to ours, it is important to note that we contend that only some funds appear to engage in WD behavior. Further, there are several factors that can help reconcile the findings of the two studies. First, our analysis is at the individual fund level rather than at the mutual fund family level, and excludes index funds and nonequity funds. Moreover, our empirical design enables us to shed light on specific fund characteristics that provide incentives to window-dress. Second, we believe that it is difficult to ascertain window dressing from institutional trade data. Thus, we use a fund's disclosed portfolio holdings and performance to develop direct measures of the propensity to window-dress. Third, we examine funds' WD behavior at fiscal quarter-ends to coincide with the timing of funds' portfolio disclosures instead of using calendar quarter-ends as in G. Hu et al. (2014).

1.2 Hypotheses

We posit that managers having low skill and achieving poor performance earlier during a quarter (e.g., during the first two months) are more likely to window-dress. The rationale is that these managers choose to window-dress as a last resort when they have performed poorly and/or have limited skill, and therefore have little expectation that they will perform better in the future. This leads to our first hypothesis:

Hypothesis 1: Window dressing is negatively related to fund performance during the first two months of a quarter and to manager skill.

A critical issue missing from the literature relates to the incentives of managers to engage in window dressing. If investors believe managers mislead by strategically changing their portfolios prior to quarter-ends, investors should punish such managers by reducing capital allocations to their funds. This poses the question: how do fund managers stand to gain by window dressing? Thus, to understand their incentives, we make two arguments. First, we contend that investors receive two signals about a manager's ability. The first signal relates to a fund's performance observed immediately at quarter-end. The second signal, relating to the portfolio's holdings, is received with a delay of up to 60 days following quarter-end. These two signals can sometimes conflict, as a fund may disclose a high (low) proportion of winner (loser) stocks but may exhibit poor performance. Such incongruence between the two signals of managerial ability can be attributable to either window dressing or stock selection. Second, we argue that a fund's performance during the delay period helps investors resolve the potential conflict between the signals. If performance is good, then investors are more likely to attribute this conflict to stock selection and reward the fund with higher incremental flows (i.e., in addition to those justified by past performance and fund characteristics). If performance is bad, then investors are more likely to attribute the conflict to window dressing and punish the fund with lower incremental flows. These two scenarios together lead to the second hypothesis:

Hypothesis 2: Relative to non–window dressers, window dressers receive higher (lower) incremental investor flows if fund performance over the reporting delay period is good (bad).

A corollary of this hypothesis is that window dressers should exhibit greater dispersion in flows across the two states of good and bad delay-period performance. This implies that window dressers are making a risky bet on their performance during the delay period where the investor flows are the payoffs of the bet.

Given the above motivations for window dressing, we consider the implications for future performance. As stated earlier, window dressers strategically alter or distort the portfolio composition with the intention to mislead investors. In addition to engaging in unnecessary portfolio churning, WD managers are unskilled and do not appear to follow information-based strategies. Together, these factors lead to our third hypothesis:

Hypothesis 3: Window dressing is negatively related to future fund performance.

2. Data and Variable Construction

We construct our data set by merging the Center for Research in Security Prices (CRSP) survivorship-bias-free mutual fund and Thomson Financial mutual fund holdings databases using the MFLINKS database from Wharton Research Data Services (WRDS). The CRSP database includes information on mutual funds' monthly returns, total net assets, inception date, fee structure, investment objectives, and other attributes. The Thomson Financial database provides quarterly or semiannual holdings of mutual funds in our sample.[3] As our focus is on U.S. equity funds, we exclude balanced, bond, international, money market, and sector funds. Since the CRSP database provides information at the share-class level, we aggregate data at the fund level by weighting each share class by its total net assets to obtain value-weighted averages of monthly returns and annual expense ratios. Our final sample comprises 59,060 quarterly reports from 2,623 equity funds that span the period September 1998 to December 2008.[4]

2.1 Measures of window dressing

We develop two measures of window dressing that are based on reported fund holdings and returns. More specifically, we propose both a relative and an absolute measure that capture the inconsistency between a fund's reported performance based on net asset values and performance imputed from its disclosed holdings.

2.1.1 *Rank Gap* Relative measure of window dressing. At the end of each fund's fiscal quarter, we create quintiles of all domestic stocks in the CRSP stock database by sorting stocks in descending order according to their returns over the past three months. The first (fifth) quintile consists of stocks that achieve the highest (lowest) returns. Then, using each fund's reported holdings, we identify stocks that belong to different quintiles and calculate the proportions of the fund's assets invested in the first and fifth quintiles. In the spirit of Lakonishok et al. (1991) and Jegadeesh and Titman (1993), we refer to these two extreme quintiles as winner and loser proportions, respectively.

Next, for each fiscal quarter that has at least 100 funds reporting holdings, we rank the funds in three ways. For the first ranking, we sort funds in descending order by their quarterly returns, with funds in the 1st percentile bin being the best performing funds (and all assigned a rank equal to 1) and funds in the 100th percentile bin being the worst (and assigned a rank equal to 100). For the second ranking, we sort funds in *descending* order according to their proportion of winner stock holdings and again assign ranks between 1 and 100, with funds in the 1st (100th) percentile bin having the highest (lowest) winner proportion. For the third ranking, we sort funds in *ascending* order according to their proportion of loser stock holdings and assign ranks similarly. Hence, funds in

[3] Under the Securities Act of 1933, the Securities Exchange Act of 1934, and the Investment Company Act of 1940, mutual fund managers are required to periodically disclose their holdings. Following a 1985 amendment, funds were required to submit annual and semiannual reports (N-CSR and N-CSRS, respectively); however, a large majority of managers voluntarily continued to disclose portfolio holdings on a quarterly basis as was previously required. Effective May 10, 2004, the SEC requires investment companies to disclose as of the end of the first and the third fiscal quarters as well, on Form N-Q. For further detail, see http://www.sec.gov/rules/final/33-8393.htm#IB.

[4] Although mutual fund holdings data are available since January 1980, our sample period starts in September 1998, corresponding to when daily fund return data became available to estimate daily four-factor alphas.

the 1st (100th) percentile bin will have the lowest (highest) loser proportion. Note that we switch the sorting order for loser stocks to make the interpretation of rankings consistent with that for the winner stocks (i.e., a high proportion of winners is analogous to a low proportion of losers). We illustrate the three percentile rankings as follows:

Rank	Fund Performance	Winner Proportion	Loser Proportion
1	1 (best performance)	1 (highest proportion)	1 (lowest proportion)
2	2	2	2
3	3	3	3
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
98	98	98	98
99	99	99	99
100	100 (worst performance)	100 (lowest proportion)	100 (highest proportion)

A well-performing fund should have a high rank based on fund performance and corresponding high ranks based on winner and loser proportions. Similarly, a poorly performing fund should have low ranks based on all three dimensions. However, a fund with a low performance rank, but relatively high rankings of winner and loser proportions, should have a greater likelihood of window dressing. We thus compute *Rank Gap* as the difference in a fund's performance rank and the average of the two ranks based on winner and loser proportions:

$$Rank\ Gap = \left\{ Performance\ Rank - \frac{WinnerRank + LoserRank}{2} \right\} / 200, \qquad (1)$$

where *Performance Rank* is the rank of the fund performance, *WinnerRank* is the rank of the winner proportion, and *LoserRank* is the rank of the loser proportion. We scale the measure by 200 to produce a theoretical bound of (−0.495, +0.495). The higher the *Rank Gap*, the greater the likelihood of window dressing. In Panel A of Table 1, we report summary statistics for the *Rank Gap* measure and observe that its mean (median) in our sample is −0.0003 (−0.0025).

2.1.2 *BHRG*: Absolute measure of window dressing. Our second measure is "backward holding return gap" (*BHRG*), defined as the difference between (a) the quarterly return (*net* of expenses and trade costs) of a hypothetical portfolio comprising a fund's end-of-quarter holdings that are assumed to have been held throughout the quarter, and (b) the fund's actual quarterly return (*Actual Return* or *AR*):[5]

$$BHRG = Backward\ Holding\ Return\ (BHR) - Actual\ Return\ (AR). \qquad (2)$$

[5] In computing the "backward holding return" or *BHR* of the hypothetical portfolio, we follow KSZ and assume a buy-and-hold strategy and make appropriate adjustments to the number of shares held and stock prices for any stock splits and other share adjustments occurring during the quarter.

Table 1
Descriptive summary statistics and correlation coefficients

Variable	*N*	Mean	Median	Max	Min
Panel A: Window-dressing measures					
BHRG	58517	0.0072	0.0018	0.1643	−0.0863
Rank Gap	57236	−0.0003	−0.0025	0.4950	−0.4525
Panel B: Other variables					
3-Month Alpha	58181	−0.0033	−0.0033	0.1424	−0.13617
Manager Skill	56433	0.0000	−0.0001	0.0118	−0.0126
Expense	58159	0.0128	0.0125	0.0290	0
TNA ($ million)	59060	1182	238	17678	2
Turnover	58906	0.1222	0.0969	0.6703	0
Load	44680	0.6838	1	1	0
Flow	58828	0.0192	−0.0106	1.1101	−0.2941
Trade Cost	58933	0.0016	0.0010	0.0125	−0.0005

Panel C: Correlations

	Rank Gap	BHRG	3-Month Alpha	Manager Skill	Expense	TNA	Turnover	Load	Flow
BHRG	0.53***								
3-Month Alpha	−0.33***	−0.09***							
Manager Skill	−0.13***	−0.18***	0.04***						
Expense	0.06***	0.06***	−0.03***	−0.00					
TNA	−0.03***	−0.02**	0.02***	−0.02***	−0.28***				
Turnover	0.14***	0.32***	0.01	0.00	0.20***	−0.13***			
Load	0.03***	−0.01	−0.03***	−0.00	0.24***	−0.02**	0.03***		
Flow	−0.11***	−0.02***	0.11***	0.00	0.03***	−0.03***	0.14***	−0.01**	
Trade Cost	0.13***	0.28***	−0.01*	0.01	0.19***	−0.14***	0.69***	0.01**	0.26***

The table reports the summary statistics and correlation coefficients of the key variables used in the empirical analysis. Our sample includes 59,060 fund-quarter observations for 2,623 funds between September 1998 and December 2008. Backward holding return gap (*BHRG*) is defined as the quarterly return of a hypothetical portfolio that is assumed to have been invested at the beginning of the quarter in the fund's disclosed end-of-quarter holdings after subtracting the expenses and trade costs (backward holding return, *BHR*), minus the fund's reported return (actual return). *Rank Gap* is defined as the percentile rank of fund performance (performance rank) over a fiscal quarter minus the average of the winner rank based on winner proportion and loser rank based on loser proportion over the same quarter. Winner (loser) proportion is the proportion of the fund's assets invested in winning (losing) stocks that achieve good (poor) performance over the quarter, where winning (losing) stocks are those that have returns in the top (bottom) 20% of all stocks in that quarter. *Rank Gap* is scaled to lie between –0.5 and 0.5, for which we divide the difference in performance rank and the average of the ranks based on winner and loser proportions by 200. *3-Month Alpha* is computed after summing the daily four-factor alphas computed each quarter using daily fund returns. *Manager Skill* is defined as the moving average of 12 monthly return gaps as in Kacperczyk, Sialm, and Zheng (2008). *Expense* is the annual expense ratio. *TNA* is the total net assets under management at the end of the quarter. *Turnover* is computed from the portfolio holdings data of mutual funds as the minimum of the total purchases and sales in a quarter divided by beginning-of-the-quarter assets. *Load* is a dummy variable defined as 1 if there is any front-end or back-end load and 0 otherwise. *Flow* is dollar fund flows over a quarter scaled by beginning-of-the-quarter assets. *Trade Cost* is the trade cost over a quarter in buying and selling stocks as a percentage of the beginning-of-the-quarter assets. *BHRG*, *3-Month Alpha*, *Manager Skill*, *Expense*, *TNA*, *Turnover*, *Flow*, and *Trade Cost* are winsorized at the 1st and 99th percentiles. *, **, and *** denote significant differences from zero at the 10%, 5%, and 1% levels, respectively.

The higher the *BHRG*, the greater the propensity to window dress. In Panel A of Table 1, we also report summary statistics for *BHRG* and show that the mean (median) is 0.0072 (0.0018). In our subsequent empirical analysis, we use both WD measures (*Rank Gap* and *BHRG*) in their continuous forms. We also construct indicator variables of the greater propensity to window-dress based on the top 10% and 20% values of the *Rank Gap* and *BHRG* continuous measures.

2.1.3 Discussion of window dressing and past performance. As described above, the two WD measures are functions of actual reported performance and performance imputed from funds' reported holdings. As such, there is a potential concern that there is a mechanical relation between the WD measures and past performance. We acknowledge that such a relation cannot be completely ruled out, as past performance, by construction, is a key component of the two measures. Thus, in the empirical analyses to follow, we explicitly control for past performance. Specifically, when we examine how both investor flows and future fund performance relate to window dressing, we include linear and nonlinear functions of funds' past performance to show that window dressing does contain predictive information beyond that contained in past performance.

2.2 Other key variables

Alpha. We estimate daily alphas based on the four-factor model of Carhart (1997) that are then summed to compute alphas of longer specified intervals, for example, one-month, two-month, three-month (i.e., quarterly), one-year, two-year, and three-year. As noted, it is important to control for the momentum effect since it shares with window dressing the feature of buying winners and selling losers. However, the widely used momentum factor returns from Ken French's Web site are not appropriate in our context for two reasons. First, we require a three-month evaluation period to match funds' reporting horizon instead of French's eleven-month evaluation period. In addition, instead of French's one-month holding period, we require a three-month holding period to help distinguish between the horizons of a momentum trader and a window dresser. Second, we follow Lakonishok et al. (1991) and Jegadeesh and Titman (1993) and define winner (loser) stocks as the top (bottom) 20% of performers during a quarter. Hence, for consistency, we compute momentum factor returns by taking long and short positions in stocks in the top and bottom 20%, respectively, instead of the top and bottom 30% as used by French.

Due to their propensity to buy winners and sell losers, window dressers may show a greater increase in exposure to or loading on the momentum factor toward the end of a quarter. To account for these potential changes and thus mitigate bias in the estimation of the daily alphas, we repeat the estimation of the four-factor model on a monthly basis.[6] Using daily fund and factor returns from a given month along with the estimated factor loadings, we obtain estimates of daily alphas. We then sum these daily alphas to obtain the alphas of various maturities. Panel B of Table 1 shows that the mean (median) quarterly alpha is −0.33% (−0.33%).

[6] We do find evidence of greater intraquarter variation in momentum betas for window dressers. We estimate monthly betas on the momentum factor for each fund over each fund quarter and compute the difference between the third and the average of the first and second month betas. We then sort these differences into two groups, window dressers and non–window dressers, and conduct univariate and multivariate tests between each group. In results reported in the Supplementary Appendix (Tables SA.1 and SA.2), we find that the difference between the third-month beta and the average of the first two months' betas is significantly higher for window dressers.

Flow. We calculate monthly net fund flows as $[TNA_t - TNA_{t-1} \cdot (1+r_t)]/TNA_{t-1}$, where TNA_t and TNA_{t-1} are the fund's total net assets under management at the end of months t and $t-1$, respectively, and r_t is the net-of-fee return during month t. Quarterly fund flows are computed by summing the dollar flows over the three months of the quarter and dividing by the total net assets at the beginning of the quarter. In Panel B of Table 1, we observe that the mean (median) quarterly flow is 1.92% (−1.06%).

Trade Cost. We obtain information for computing trade costs from daily institutional trades reported in the Abel Noser database during our sample period. To compute a fund's trade costs in a given fund quarter, we identify the fund's buys and sells of each stock traded during the quarter by comparing its beginning and ending holdings. Then, for each stock bought or sold, we identify in the Abel Noser database all institutions' buys and sells of that stock during the same quarter.[7] For each trade (keeping buys and sells separate) we compute the explicit trade cost per share by dividing the reported trade commission by the number of shares in the transaction. We compute the implicit trade cost per share for buys (sells) as the difference between the reported trade price (open price) and the stock's opening price (trade price) that day. The total trade cost per share is computed as the sum of the explicit and implicit trade costs. Then, for all trades of the stock during the quarter, we compute separately for buys and sells the volume-weighted explicit, implicit, and total (unit) trade cost per share. We repeat these calculations for each stock traded by the fund in the quarter.

We next link these unit trade costs to a fund's buys and sells during the quarter and compute, for each trade, the explicit, implicit, and total trade costs by multiplying the unit trade costs by the trade size. We then sum the trade costs for all trades by the fund during the quarter to obtain the total dollar trade costs, which we scale by the fund's beginning-of-quarter *TNA*. We report in Panel B of Table 1 that the mean (median) quarterly trade cost is 0.16% (0.10%).

Turnover. We compute a fund's quarterly turnover ratio as the minimum of the dollar values of purchases and sales, divided by total net assets at the beginning of the quarter. In Panel B of Table 1, we report the mean (median) quarterly portfolio turnover to be 12.2% (9.7%).

Manager Skill. For manager skill, we follow KSZ and use the 12-month moving average of the monthly return gap, which they show is positively related to future performance. In Panel B of Table 1, we report that the mean and median manager skill are both close to zero.

Style. We use the investment objective code (IOC) field from the Thomson Financial mutual fund holdings database to construct style dummies. We exclude the four nonequity styles (international, municipal bonds, balanced,

[7] We note that individual institutions are not identified in the Abel Noser database. Hence, we compute trade costs for each stock based on the trades of all institutions during a given quarter of that stock.

and bonds & preferreds) and focus on the five active equity styles: aggressive growth, growth, growth & income, metals, and unclassified.[8]

Other variables used in the analysis include *Size*, defined as the log of a fund's *TNA*; *Expense*, defined as a fund's annual expense ratio; and *Load*, defined as an indicator variable having a value of 1 if a fund has a front-end or back-end load, and 0 otherwise.

2.3 Correlations

Panel C of Table 1 provides the correlations between the key variables. The two WD measures, *Rank Gap* and *BHRG*, have a strong positive correlation of 0.53. In addition, we observe a negative correlation between both measures and quarterly alpha (−0.33 with *Rank Gap*, and −0.09 with *BHRG*). Further, the two measures are negatively correlated with manager skill (−0.13 and −0.18, respectively). Although these correlations are based on contemporaneous values and therefore do not necessarily imply causality, it is interesting that the signs of the correlations are consistent with our first hypothesis suggesting that WD behavior is negatively related to past performance and manager skill. We also find that the WD measures are positively related to trade costs, expense ratio, and turnover, and negatively related to flows.

3. Window Dressing: Motivation and Attributes

3.1 Do investors respond to portfolio characteristics?

In addition to information contained in past performance, there is growing evidence in the academic literature that investors use information based on disclosed portfolio holdings to assess managerial ability.[9] This evidence does not necessarily imply irrationality, since rational investors can use holdings information as an ex ante measure of managerial ability in conjunction with past performance, which is an ex post measure. If this is indeed the case then investor flows should respond to portfolio characteristics after controlling for past performance. In the context of our study, these characteristics relate to the proportions of winners and losers in the disclosed portfolios. We examine the relation between fund flows and proportions of winners and losers, controlling for past performance and other fund characteristics, and estimate the following

[8] If a fund's IOC is unclassified, we use the Lipper objective codes (EIEI, G, LCCE, LCGE, LCVE, MCCE, MCGE, MCVE, MLCE, MLGE, MLVE, SCCE, SCGE, SCVE), the Strategic Insight objective codes (AGG, GMC, GRI, GRO, ING, SCG), and Wiesenberger Fund Type codes (G, G-I, AGG, GCI, GRI, GRO, LTG, MCG, SCG) to identify whether the fund is an actively managed equity fund for inclusion in our sample.

[9] See, for example, Grinblatt and Titman (1989, 1993); Grinblatt, Titman, and Wermers (1995); Daniel et al. (1997); Wermers (1999, 2000); Chen, Jegadeesh, and Wermers (2000); Gompers and Metrick (2001); Cohen, Coval, and Pastor (2005); Kacperczyk, Sialm, and Zheng (2005, 2008); Sias, Starks, and Titman (2006); Alexander, Cici, and Gibson (2007); Jiang, Yao, and Yu (2007); Kacperczyk and Seru (2007); Cremers and Petajisto (2009); Baker et al. (2010); Huang and Kale (2013).

regression:

$$Flows_{i,t+1} = \beta_0 + \beta_1 Winner\ Prop_{i,t} + \beta_2 Loser\ Prop_{i,t} + \beta_3 Alpha_{i,t}$$
$$+ \beta_4 Manager\ Skill_{i,t-1} + \beta_5 Expense_{i,t} + \beta_6 Size_{i,t} + \beta_7 Turnover_{i,t} + \beta_8 Load_{i,t}$$
$$+ \beta_9 Trade\ Cost_{i,t} + Style\ dummies + Time\ dummies + \psi_{i,t} \quad (3)$$

where $Flows_{i,t+1}$ is the quarterly percentage net flow for fund i in quarter $t+1$, $Winner\ Prop_{i,t}$ ($Loser\ Prop_{i,t}$) is the proportion of assets of fund i invested in the top (bottom) quintile of stocks in quarter t, $Alpha_{i,t}$ is the alpha of fund i over quarter t, $\psi_{i,t}$ is the error term, and all other variables are as defined in Section 2.2.[10] We also estimate an alternative specification that allows for nonlinearity in the relation between flows and performance. Following Sirri and Tufano (1998), we use a piecewise linear specification where a fund's performance each quarter is classified in one of three performance groups including the top and bottom quintiles and the middle 60%. Specifically, we replace $Alpha_{i,t}$ in Equation (3) with $QtrAlphaTop_{i,t}$, $QtrAlphaMid_{i,t}$, and $QtrAlphaBot_{i,t}$, defined as the top 20%, middle 60%, and bottom 20% performance quintiles for fund i in quarter t. In all of our empirical tests, we cluster standard errors by fund and include fixed effects for time and funds' investment styles.

Table 2 reports the results from the regression estimations. The findings support the notion that investors respond to portfolio characteristics over and above a fund's past performance. From Column 1 for the linear performance specification, we observe a positive and significant estimated coefficient on the winner proportion (coeff. = 0.0560, p-value = 0.000), and a negative and significant estimated coefficient on the loser proportion (coeff. = −0.0983, p-value = 0.000). It is important to note that the observed significant relation between fund flows and certain portfolio characteristics (i.e., winner and loser proportions) is in addition to the flows being driven by past performance (coeff. = 0.3353, p-value = 0.000), as has been documented in the extant literature (e.g., Chevalier and Ellison 1997; Sirri and Tufano 1998). We observe similar findings in Column 2 where we allow for a nonlinear relation between flows and performance. Finally, we observe a positive relation between fund flows and manager skill, expense ratio, and trade costs and a negative relation with portfolio turnover and load.

3.2 Determinants of window dressing

Our first hypothesis is that window dressing is associated with unskilled managers and funds performing poorly during the first two months of a quarter. We test this hypothesis using conditional double sorts on skill and performance

[10] We use alphas rather than the actual (raw) returns because winner and loser proportions are likely to be highly correlated with actual returns.

Table 2
Quarterly fund flows and proportion of winners and losers

Dependent variable: Flows during quarter $t+1$

Variables	(1)	(2)
WinnerProp$_t$	0.0560***	0.0535***
	0.000	0.000
LoserProp$_t$	−0.0983***	−0.0958***
	0.000	0.000
Quarter Alpha$_t$	0.3353***	
	0.000	
QtrAlphaBot$_t$		0.0657***
		0.000
QtrAlphaMid$_t$		0.0329***
		0.000
QtrAlphaTop$_t$		0.1265***
		0.000
Manager Skill$_{t-1}$	1.6792***	1.6941***
	0.000	0.000
Expense$_t$	0.7939**	0.7390**
	0.013	0.020
Size$_t$	0.0323	0.0390
	0.651	0.583
Turnover$_t$	−0.0286*	−0.0255
	0.095	0.137
Load$_t$	−0.0165***	−0.0160***
	0.000	0.000
Trade Cost$_t$	0.0375***	0.0376***
	0.001	0.001
Constant	−0.0390**	−0.0624***
	0.017	0.000
Time and Style dummies	Yes	Yes
Observations	29,467	29,467
Adj. R^2	0.0531	0.0539

This table reports the results of regressions using quarterly percentage net fund flows during the lead quarter as the dependent variable. Independent variables include *WinnerProp* and *LoserProp*, the proportion of funds' assets invested in the top and bottom return quintiles of stocks during a quarter, respectively. $QtrAlphaTop_t$, $QtrAlphaMid_t$, and $QtrAlphaBot_t$ are the top 20%, middle 60%, and bottom 20% performance quintiles for a fund in quarter t as defined in Sirri and Tufano (1998). *Size* is the natural logarithm of total net assets (*TNA*) at quarter-end. Other variables are as defined in Table 1. Standard errors are adjusted for clustering at the fund level. p-values are reported below the estimated coefficients. *, **, and *** denote significant differences from zero at the 10%, 5%, and 1% levels, respectively.

as well as multivariate regressions. An advantage of the sorting method is that it does not impose linearity on the relation between window dressing and either skill or performance.

In Table 3, we present the results of our sorting analysis, where we first sort funds into manager skill quintiles and then, within each skill quintile, sort funds into performance (*2-Month Alpha*) quintiles. Panels A and B report the averages of *Rank Gap* and *BHRG* for the twenty-five double-sorted portfolios. In both panels, controlling for managerial skill, as we move by rows from the lowest to highest performance quintile, the average WD measure is generally monotonically decreasing. Similarly, controlling for performance, as we move by columns from the lowest to highest skill quintile, the average WD measure again is generally monotonically decreasing. Also, the (5–1) differences for both the extreme performance and skill quintiles are all highly significant. Further, we can observe the interaction effects of skill and performance on window

Table 3
Prior performance and window dressing: Results from double 5 × 5 sorts

Panel A: Means of *Rank Gap*

		2-Month Alpha					
		1 (low)	2	3	4	5 (high)	5–1
	1 (low)	0.0861	0.0495	0.0292	−0.0083	−0.0301	−0.1162
		0.00	0.00	0.00	0.00	0.00	0.00
	2	0.0367	0.0223	0.0041	−0.0213	−0.0417	−0.0784
		0.00	0.00	0.05	0.00	0.00	0.00
Manager Skill	3	0.0361	0.0156	−0.0070	−0.0253	−0.0422	−0.0783
		0.00	0.00	0.00	0.00	0.00	0.00
	4	0.0321	0.0143	−0.0056	−0.0284	−0.0438	−0.0759
		0.00	0.00	0.00	0.00	0.00	0.00
	5 (high)	0.0409	0.0120	−0.0154	−0.0441	−0.0653	−0.1062
		0.00	0.00	0.00	0.00	0.00	0.00
	5–1	−0.0452	−0.0375	−0.0446	−0.0358	−0.0353	
		0.00	0.00	0.00	0.00	0.00	

Panel B: Means of *BHRG*

	1 (low)	0.0322	0.0162	0.0128	0.0116	0.0177	−0.0145
		0.00	0.00	0.00	0.00	0.00	0.00
	2	0.0112	0.0054	0.0054	0.0038	0.0038	−0.0073
		0.00	0.00	0.00	0.00	0.00	0.00
Manager Skill	3	0.0096	0.0044	0.0020	0.0009	0.0018	−0.0078
		0.00	0.00	0.00	0.03	0.01	0.00
	4	0.0086	0.0039	0.0022	0.0023	0.0008	−0.0078
		0.00	0.00	0.00	0.00	0.33	0.00
	5 (high)	0.0127	0.0051	0.0038	0.0023	0.0006	−0.0121
		0.00	0.00	0.00	0.03	0.70	0.00
	5–1	−0.0195	−0.0111	−0.0090	−0.0092	−0.0171	
		0.00	0.00	0.00	0.00	0.00	

This table reports means of *Rank Gap* (in Panel A) and *BHRG* (in Panel B) for twenty-five portfolios of mutual funds sorted first by their manager skill measure and then by their alphas measured over the first two months of the quarter. The *2-Month Alpha* is the sum of daily four-factor alphas over the first two months in a quarter. All other variables are as defined in Table 1. p-values of the t-tests are reported below the means after adjusting the standard errors for clustering at the fund level.

dressing. In Panel A, we find that (a) the highest and lowest mean values of *Rank Gap* are in cells (1,1) and (5,5) with values of 0.0861 and −0.0653, respectively; and (b) the values decrease monotonically along this diagonal. We observe a similar pattern in Panel B for *BHRG*. Together these findings support Hypothesis 1 that window dressing is negatively related to manager skill and funds' first two months' performance during the quarter.

We next estimate OLS and logistic regressions of the following form:

$$\begin{aligned} WD_{i,t} = \lambda_0 + \lambda_1 \textit{Two-month Alpha}_{i,t} + \lambda_2 \textit{Manager Skill}_{i,t-1} + \lambda_3 \textit{Expense}_{i,t} \\ + \lambda_4 \textit{Turnover}_{i,t} + \lambda_5 \textit{Size}_{i,t} + \lambda_6 \textit{Load}_{i,t} + \lambda_7 \textit{Trade Cost}_{i,t} \\ + \textit{Style dummies} + \textit{Time dummies} + \xi_{i,t} \end{aligned} \quad (4)$$

where $WD_{i,t}$ is the WD measure for fund i in quarter t, specified as a continuous (indicator) variable in the OLS (logistic) specification and as defined in Section 2.1.2; $\xi_{i,t}$ is the error term; and the other variables are as defined previously in Section 2.2.

Table 4
Determinants of window dressing: Multivariate analysis

Dependent variable: Window dressing during quarter *t*

Variables	*BHRG*	*BHRG10% Dummy*	*BHRG20% Dummy*	*Rank Gap*	*Rank Gap10% Dummy*	*Rank Gap20% Dummy*
2-Month Alpha$_t$	−0.0902***	−6.2419***	−4.3338***	−0.8738***	−16.3321***	−15.3831***
	0.000	0.000	0.000	0.000	0.000	0.000
Manager Skill$_{t-1}$	−0.7528***	−39.3524***	−29.6760***	−2.3133***	−39.1142***	−33.5428***
	0.000	0.000	0.000	0.000	0.000	0.000
Expense$_t$	0.1033	76.2404***	49.8678***	0.4577	45.3936***	36.5311***
	0.488	0.000	0.000	0.131	0.000	0.000
Size$_t$	0.1424***	12.2076***	11.3944***	−0.0170	1.7436	1.4733
	0.000	0.002	0.000	0.825	0.505	0.415
Turnover$_t$	0.0688***	6.2330***	5.8776***	0.0735***	3.1782***	2.1908***
	0.000	0.000	0.000	0.000	0.000	0.000
Load$_t$	−0.0016	−0.3879***	−0.1324	0.0051**	−0.0709	0.0224
	0.237	0.003	0.127	0.048	0.362	0.680
Trade Cost$_t$	0.0341***	1.6859***	1.8790***	0.0502***	1.0775***	0.9396***
	0.000	0.000	0.000	0.000	0.000	0.000
Constant	−0.0099***	−4.5955***	−3.3761***	−0.0239***	−3.6541***	−2.5127***
	0.008	0.000	0.000	0.002	0.000	0.000
Time & Style dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30,874	30,874	30,874	30,072	30,072	30,072
Adj. R^2	0.172			0.103		
Pseudo R^2		0.178	0.136		0.0916	0.0677

This table reports the results of regressions of the two WD measures on fund characteristics. *Rank Gap10% (20%) Dummy* is an indicator variable defined as 1 if *Rank Gap* is in the top 10th (20th) percentile for a given quarter, and 0 otherwise; and *BHRG10% (20%) Dummy* is an indicator variable defined as 1 if *BHRG* is in the top 10th (20th) percentile for a given quarter, and 0 otherwise. All other variables are as defined in Tables 1, 2, and 3. Standard errors are adjusted for clustering at the fund level. *p*-values are reported below the estimated coefficients. *, **, and *** denote significant differences from zero at the 10%, 5%, and 1% levels, respectively.

Table 4 reports the results from OLS and logistic regressions. The estimated coefficients of the performance and manager skill variables are negative and significant at the 1% level in all specifications, confirming our findings from the double-sort analysis. For example, using the continuous form of the *BHRG* measure (see Column 2), the estimated coefficient on *2-month Alpha* is −0.0902, and on manager skill it is −0.7528. Similarly, when using the continuous form of the *Rank Gap* measure as the dependent variable (see Column 5), the corresponding estimated coefficients are −0.8738 and −2.3133, respectively. For the regression based on the indicator variable representing the top 10% values of *BHRG* (see Column 3), we find that the estimated coefficients on alpha and skill are −6.2419 and −39.3524, respectively. For the top 10% values of *Rank Gap* (see Column 6), the estimated coefficients on alpha and skill are −16.3321 and −39.1142, respectively. We find similar results for the top 20% indicator variable specifications (see Columns 4 and 7). Taken together, these findings support our first hypothesis that unskilled managers and funds that have performed poorly earlier in the quarter are more likely to exhibit WD behavior. We also find that the estimated coefficients on turnover and trade costs are positive and highly significant in all specifications, indicating that these variables are strongly associated with WD behavior. We further investigate these associations in the following sections.

3.3 The dynamics of window-dressing turnover

As window dressing is not information-based, its associated trading should result in unnecessary trade costs and lower performance. However, quantifying and comparing window dressing to "non-window-dressing" turnover presents at least two challenges. The first challenge involves parsing a fund's trading activity into its WD and non-WD components. At first glance, these two components appear to be independent, as window dressing should be related to buying winners and selling losers, and non-WD activity should include all remaining trading. However, a WD strategy requires financing. The strategy could be self-financing if the proceeds from selling losers are sufficient to buy winners. Otherwise, assuming a fund's cash holdings are held constant, there are two other potential sources for filling the deficit: (a) fund inflows (if any) and, if these are insufficient, (b) the selling of nonloser stocks. This latter component, selling nonloser stocks in order to finance a WD strategy, can be argued to be part of WD turnover rather than non-WD turnover. We can thus think of WD activity (*WDA*) as having a direct component equal to the dollar value of buying winners and selling losers, plus a potential indirect component (*IndWDA*) related to the selling of other stocks of an amount necessary to complete the financing of the purchase of winners. We refer to the sum of these two components as adjusted *WD* activity (*AdjWDA*). We express these relations as follows:

$$AdjWDA = WDA + IndWDA, \tag{5}$$

$$IndWDA = Max\{(Buy\ Winners - Sell\ Losers) - Max[Flows, 0], 0\}. \tag{6}$$

Next, we refer to that portion of a fund's total trading activity (*TotalTA*) that is unrelated to window dressing as non-WD activity (*NWDA*). *NWDA* activity reflects the dollar value of nonwinners bought (including losers bought) plus the dollar value of nonlosers sold (including winners sold). However, if *IndWDA* is positive, then this amount of selling of other stocks, which is a consequence of window dressing, should be netted from *NWDA* to produce an adjusted amount of non-window-dressing-related trading activity (*AdjNWDA*). We express these additional relations as follows:

$$TotalTA = All\ Buys + All\ Sells, \tag{7}$$

$$NWDA = TotalTA - WDA, \text{and} \tag{8}$$

$$AdjNWDA = NWDA - IndWDA. \tag{9}$$

A second challenge when comparing WD with non-WD activity is how to deal with the fact that their population bases are significantly different. Recall that winners and losers are defined to be 20% of their respective samples while nonwinners and nonlosers constitute the other 80%. Thus, a fund is expected to have, on average, activity in nonwinner and nonloser stocks that is approximately four times as large as its activity in winner and loser stocks. To address this issue, for each fund quarter we compute the following four turnover ratios, wherein the various trading activity measures are scaled by the fund's *TNA*: *WDA/TNA* and *NWDA/TNA*, and *AdjWDA/TNA* and *AdjNWDA/TNA*. The first pair of ratios ignores the indirect WD effect, while the latter pair of ratios accounts for this effect. We posit that the portion of overall turnover attributable to buying winners and selling losers should be relatively higher for window dressers than for non–window dressers. To test this conjecture, we divide the respective pairs of ratios to get *WDA/NWDA* and *AdjWDA/AdjNWDA* and argue that these ratios should be the highest for window dressers.

We report the findings in Table 5. In Panel A, we sort funds on *BHRG* and report the averages of the turnover measures for each decile. We observe that the direct measure of WD activity (*WDA*) is the highest for funds in decile 10 (19.1%). Moreover, we observe that the indirect WD activity turnover (*IWDA*) is also the highest in decile 10 (8.1%). This indirect activity is an economically large component of turnover, and suggests that window dressers do not solely rely on selling losers to buy winners, but also on selling substantial amounts of other stocks. We also report averages for non-window-dressing-related turnover (see columns *NWDA* and *AdjNWDA*), which are significantly larger than those for WD-related turnover (by a factor of about four for the entire sample). We also observe that the non-WD-related turnover is the highest in decile 10, suggesting that window dressers also have a higher turnover in other stocks. Still, as indicated in the column *WDA/NWDA*, where we take the ratio of window dressing to non-window-dressing-related turnover, we observe that relative WD activity remains the largest in decile 10. Thus, while window dressers appear to have higher rates of turnover reflecting their information-less trading, they engage in even higher relative rates of turnover of winner and loser stocks. This effect is even more pronounced when we adjust for indirect WD activity, as shown in the last column. We find similar results in Panel B for sorts using *Rank Gap*.

3.4 Window dressing and trade costs

The above finding that window dressing is positively related to turnover suggests that it should also be associated with higher trade costs. To investigate the magnitude of these costs, we compute for each fund each quarter the trade costs (explicit, implicit, and total) associated with the buying of winners, the selling of losers, and the sum of the two activities. We then sort funds into WD deciles and compute averages of the various trade costs for each decile. These results are presented in Panels A and B of Table 6 for *BHRG* and *Rank Gap*, respectively.

In each panel, the highest level of total trade costs is observed for funds in the highest WD decile 10. Under column heading "Buy Winner + Sell Loser

Table 5
Window-dressing trade activity

Panel A

BHRG decile	*WDA*	*NWDA*	$\frac{WDA}{NWDA}$	*IndWDA*	*AdjWDA*	*AdjNWDA*	$\frac{AdjWDA}{AdjNWDA}$
1 (low)	0.074	0.380	0.198	0.010	0.086	0.368	0.250
2	0.057	0.304	0.184	0.010	0.069	0.293	0.242
3	0.047	0.256	0.185	0.008	0.056	0.247	0.247
4	0.044	0.220	0.197	0.008	0.052	0.211	0.266
5	0.045	0.217	0.208	0.010	0.055	0.207	0.297
6	0.050	0.228	0.232	0.011	0.061	0.217	0.337
7	0.065	0.259	0.267	0.015	0.080	0.243	0.395
8	0.083	0.300	0.308	0.021	0.105	0.278	0.476
9	0.112	0.347	0.356	0.031	0.144	0.315	0.563
10 (high)	0.191	0.452	0.460	0.081	0.275	0.367	0.935
10–1	0.117	0.073	0.262	0.071	0.188	0.000	0.685
p-value	0.00	0.00	0.00	0.00	0.00	0.98	0.00

Panel B

Rank Gap decile	*WDA*	*NWDA*	$\frac{WDA}{NWDA}$	*IndWDA*	*AdjWDA*	*AdjNWDA*	$\frac{AdjWDA}{AdjNWDA}$
1 (low)	0.079	0.343	0.231	0.016	0.097	0.326	0.338
2	0.072	0.290	0.247	0.018	0.091	0.271	0.378
3	0.070	0.278	0.248	0.017	0.088	0.260	0.382
4	0.067	0.267	0.252	0.018	0.086	0.248	0.390
5	0.066	0.262	0.249	0.018	0.085	0.244	0.392
6	0.067	0.266	0.250	0.019	0.087	0.247	0.387
7	0.066	0.276	0.240	0.016	0.083	0.259	0.345
8	0.070	0.284	0.257	0.016	0.086	0.268	0.363
9	0.082	0.313	0.275	0.020	0.104	0.291	0.409
10 (high)	0.129	0.392	0.343	0.046	0.176	0.345	0.605
10–1	0.050	0.049	0.112	0.029	0.079	0.020	0.266
p-value	0.00	0.00	0.00	0.00	0.00	0.013	0.00

This table reports the means of components of WD trade activity for decile portfolios of mutual funds sorted by the two WD measures (*BHRG* in Panel A and *Rank Gap* in Panel B) for each quarter. *WDA* is a direct component of WD activity given by the dollar value of buying winners and selling losers. *NWDA* is non-WD activity given by the dollar value of nonwinners bought (including losers bought) plus the dollar value of nonlosers sold (including winners sold). *IndWDA* refers to an indirect component of WD activity given by the sell value of other stocks necessary to complete the financing of the purchase of winners. The sum of *WDA* and *IndWDA* is defined as adjusted WD activity (*AdjWDA*). *AdjNWDA* is defined as *NWDA* subtracted by *IndWDA*. All trade activity variables are scaled by beginning-of-the-quarter assets. p-values of the t-tests of the differences for deciles 10 and 1 are reported in the last row of the two panels after adjusting the standard errors for clustering at the fund level.

Cost," the estimates are 14.4 and 10.3 basis points for *BHRG* and *Rank Gap*, respectively. We note that these are quarterly trade costs and relate only to the buying of winners and selling of losers, and do not include the costs related to indirect WD activity. Also, when we examine the two components of trade cost, we find that higher implicit costs appear to drive these results more than explicit costs. Since the Abel Noser database does not provide the identities of the trading institutions, we cannot conclusively attribute the higher implicit costs to the urgency of window dressers to trade (buy winners and sell losers) near quarter-ends. Nevertheless, if one assumes that there is a predominance of window dressers in the aggregate trading of winners and losers, then this can be a possibility. Overall, this evidence suggests that window dressing is associated with higher trade costs.

Table 6
Window-dressing trade costs

Panel A

	Buy winner cost			Sell loser cost			Buy winner + sell loser cost		
BHRG decile	Total	Explicit	Implicit	Total	Explicit	Implicit	Total	Explicit	Implicit
1 (low)	0.00035	0.00007	0.00027	0.00028	0.00006	0.00022	0.00063	0.00013	0.00049
2	0.00027	0.00005	0.00022	0.00020	0.00004	0.00016	0.00047	0.00009	0.00038
3	0.00021	0.00004	0.00017	0.00017	0.00003	0.00014	0.00038	0.00007	0.00031
4	0.00020	0.00003	0.00016	0.00015	0.00003	0.00012	0.00034	0.00006	0.00028
5	0.00021	0.00004	0.00018	0.00014	0.00003	0.00011	0.00035	0.00006	0.00029
6	0.00023	0.00004	0.00019	0.00014	0.00003	0.00012	0.00038	0.00007	0.00031
7	0.00030	0.00005	0.00025	0.00018	0.00003	0.00015	0.00048	0.00008	0.00040
8	0.00041	0.00007	0.00034	0.00023	0.00004	0.00019	0.00064	0.00010	0.00053
9	0.00058	0.00009	0.00049	0.00027	0.00004	0.00023	0.00085	0.00013	0.00071
10 (high)	0.00113	0.00018	0.00095	0.00031	0.00005	0.00026	0.00144	0.00023	0.00120
10–1	0.00079	0.00011	0.00067	0.00003	−0.00001	0.00004	0.00081	0.00010	0.0007
p-value	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00

Panel B

	Buy winner cost			Sell loser cost			Buy winner + sell loser cost		
Rank Gap decile	Total	Explicit	Implicit	Total	Explicit	Implicit	Total	Explicit	Implicit
1 (low)	0.00040	0.00007	0.00032	0.00022	0.00004	0.00018	0.00062	0.00012	0.00050
2	0.00037	0.00007	0.00030	0.00017	0.00003	0.00014	0.00054	0.00010	0.00044
3	0.00037	0.00006	0.00030	0.00016	0.00003	0.00013	0.00052	0.00009	0.00043
4	0.00034	0.00006	0.00028	0.00016	0.00003	0.00013	0.00050	0.00009	0.00041
5	0.00033	0.00005	0.00027	0.00015	0.00003	0.00012	0.00048	0.00008	0.00040
6	0.00035	0.00006	0.00029	0.00019	0.00003	0.00015	0.00053	0.00009	0.00044
7	0.00030	0.00005	0.00025	0.00021	0.00004	0.00017	0.00051	0.00009	0.00042
8	0.00032	0.00005	0.00027	0.00023	0.00004	0.00019	0.00055	0.00009	0.00045
9	0.00039	0.00007	0.00032	0.00027	0.00005	0.00023	0.00066	0.00011	0.00055
10 (high)	0.00071	0.00011	0.00059	0.00032	0.00005	0.00027	0.00103	0.00017	0.00086
10–1	0.00031	0.00004	0.00027	0.00010	0.00001	0.00009	0.00042	0.00005	0.00037
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

This table reports average trade costs associated with the buying of winners and selling of losers during a quarter for decile portfolios of mutual funds sorted by the two WD measures (*BHRG* in Panel A and *Rank Gap* in Panel B) for each quarter. To compute funds' trade costs in a given fund quarter, we first compute a fund's buys and sells of each stock traded during the quarter by comparing the fund's beginning- and end-of-quarter holdings. We then use the Abel Noser database to identify all institutions' buys and sells of each stock in the same matching quarter. For each trade (keeping buys and sells separate), we compute the explicit trade cost per share by dividing the reported trade commission by the number of shares in the transaction. We also compute the implicit trade cost per share for buys (sells) as the difference between the reported transaction price (open price) and the stock's opening price (transaction price) that day. The total trade cost per share is computed as the sum of the explicit and implicit trade costs. Then, for all trades of the stock during the quarter, we compute separately for buys and sells the volume-weighted explicit, implicit, and total trade costs per share. We repeat these calculations for each stock traded by funds in the quarter. Following this, we then link these trade costs to a fund's buys and sells during the quarter and compute for each trade the explicit, implicit, and total trade costs by multiplying the trade costs per share by the trade size. We then sum the trade costs for all trades during quarter to obtain the total dollar trade costs, which we then scale by the funds' beginning-of-quarter assets under management. p-values of the t-tests of the differences for deciles 10 and 1 are reported in the last row after adjusting the standard errors for clustering at the fund level.

4. Window Dressing versus Momentum Trading

While window dressing and momentum trading share similar traits in that both involve buying winners and selling losers, our findings so far support that our measures are capturing window dressing, as they are negatively correlated with fund's past performance and manager skill (Hypothesis 1). We conduct two types of seasonality tests, intraquarter and December, to distinguish between window dressing and information-motivated momentum trading.

4.1 Intraquarter tests

We examine the intraquarter exposure of a fund's returns to its reported winner stocks. For momentum traders, this exposure should be uniformly distributed across months due to the monthly updating of winner stocks inherent in the strategy. That is, a momentum trader should buy winners and sell losers on a rolling basis. Consider, for example, a momentum trader who discloses holdings on the March quarterly cycle. Given the holding period of a typical momentum trader, winners acquired in January, February, and March will have a strong likelihood of still being held and thus reported at March end. For a window dresser, winner stocks are only acquired periodically at fiscal quarter month-ends, for example, in March. Thus, window dressers' exposure to winners should be systematically higher in the third month of a quarter.[11]

To test this conjecture, we estimate this exposure for each fund month by computing the correlation over the month between a fund's daily returns and the daily returns on the portfolio of recent winner stocks held by the fund. Recent winners are determined over the three-month period ending each month and are updated on a monthly basis. To determine if winners are actually held by a fund at month-end, we use the fund's reported quarter-end holdings to proxy for month-end holdings. We believe that any noise introduced by this proxy should bias us against finding results consistent with our conjecture. We then estimate the following regression:

$$\begin{aligned} Corr_{i,t} = {} & \alpha_0 + \alpha_1 WD_{i,t} + \alpha_2 WD_{i,t} \times I(FQEM)_{i,t} + \alpha_3 I(FQEM)_{i,t} \\ & + Style\ dummies + \eta_{i,t} \end{aligned} \tag{10}$$

where $Corr_{i,t}$ is the correlation between daily fund returns and daily returns of winner stocks held by the fund i in month t; $I(FQEM)_{i,t}$ is an indicator variable that takes a value of 1 for fund i if month t is a fiscal quarter-ending month, and 0 otherwise; and $\eta_{i,t}$ is the error term. For window dressers, unlike momentum traders, the monthly correlation should increase systematically during the month corresponding to the fiscal quarter-end, that is, α_2 should be greater than zero. From results in Table 7, we observe that the estimated coefficient α_2 is positive and significant in all six specifications, except in model 1, where it is positive, but not significant.

4.2 December tests

The literature on tournaments and the flow-performance relation (e.g., see Brown, Harlow, and Starks 1996; Chevalier and Ellison 1997; Sirri and Tufano

[11] We thank Clemens Sialm for directing us to this conjecture.

Table 7
Intraquarter variability in fund exposure to winner stocks

Dependent variable: Monthly correlation between fund returns and returns on portfolio of recent winners

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Rank Gap	−0.0637***					
	0.000					
Rank Gap × *I(FQEM)*	0.0016					
	0.856					
Rank Gap10% Dummy		−0.0181***				
		0.002				
Rank Gap10% Dummy × *I(FQEM)*		0.0097***				
		0.001				
Rank Gap20% Dummy			−0.0192***			
			0.000			
Rank Gap20% Dummy × *I(FQEM)*			0.0058***			
			0.008			
BHRG				0.5944***		
				0.000		
BHRG x *I(FQEM)*				0.1369***		
				0.000		
BHRG10% Dummy					0.0401***	
					0.000	
BHRG10% Dummy × *I(FQEM)*					0.0155***	
					0.000	
BHRG10% Dummy						0.0406***
						0.000
BHRG10% Dummy × *I(FQEM)*						0.0148***
						0.000
I(FQEM)	0.0105***	0.0095***	0.0094***	0.0093***	0.0088***	0.0073***
	0.000	0.000	0.000	0.000	0.000	0.000
Constant	0.8515***	0.8532***	0.8555***	0.8354***	0.8386***	0.8323***
	0.000	0.000	0.000	0.000	0.000	0.000
Style dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	115,619	115,619	115,619	119,113	119,113	119,113
Adj. R^2	0.0171	0.0164	0.0172	0.0309	0.0198	0.0245

This table reports the results of regressions of the monthly correlations between daily fund returns and the daily returns on the portfolio of recent winner stocks held by the fund. *BHRG10% Dummy* and *Rank Gap10% Dummy* are indicator variables that take a value of 1 if the fund is in the top *BHRG* and *Rank Gap* decile in a month, and 0 otherwise. *I(FQEM)* is an indicator variable that takes a value of 1 in a month that is a fund's fiscal-quarter-end month, and 0 otherwise. Other variables are as defined in Tables 1 and 4. Standard errors are adjusted for clustering at the fund level. p-values are reported below the estimated coefficients. *, **, and *** denote significant differences from zero at the 10%, 5%, and 1% levels, respectively.

1998; Huang, Wei, and Yan 2007; Hu et al. 2011) suggests that investors evaluate funds on a calendar-year basis, which may provide greater incentives to window-dress in December. Also, window dressers who sell losing stocks in December might disguise their behavior by pooling themselves with tax-loss sellers. Moskowitz (2000) alludes to this possibility that window dressing and tax-motivated trading may be concentrated in December. We therefore conjecture that window dressing should be more pronounced in December quarter-end months compared with other months, while momentum trading should be more uniformly distributed over the calendar year.

We conduct two December-related tests. First, we modify and reestimate Equation (4) after including a December-end quarter indicator variable as an additional independent variable. We use only *BHRG* as the dependent variable and note that we cannot use the *Rank Gap* measure or any of the indicator

Table 8
Seasonality in the intraquarter variability of the *BHRG* measure

	1st – 3rd month *BHRG*		Avg.(1st + 2nd) – 3rd month *BHRG*	
Variables	(1)	(2)	(3)	(4)
BHRG10% Dummy	0.0128***		0.0095***	
	0.000		0.000	
BHRG10% Dummy x *Dec*	0.0068***		0.0048***	
	0.000		0.000	
BHRG20% Dummy		0.0104***		0.0076***
		0.000		0.000
BHRG20% Dummy x *Dec*		0.0057***		0.0040***
		0.000		0.000
Dec	0.0055***	0.0053***	0.0104***	0.0103***
	0.000	0.000	0.000	0.000
Constant	−0.0103***	−0.0114***	−0.0147***	−0.0156***
	0.000	0.000	0.000	0.000
Time & Style dummies	Yes	Yes	Yes	Yes
Observations	40,695	40,695	40,695	40,695
Adj. R^2	0.137	0.152	0.147	0.157

This table reports the results of regressions of the intraquarter variation in the *BHRG* measure. In models (1) and (2) the dependent variable is the third-month *BHRG* subtracted from the first-month *BHRG* in each quarter, while in models (3) and (4) the dependent variable is the third-month *BHRG* subtracted from the average monthly *BHRG* over the first two months. *Dec* is an indicator variable that equals 1 if the reporting month is December, and 0 otherwise. Other variables are as defined in Tables 1 and 4. Standard errors are adjusted for clustering at the fund level. p-values are reported below the estimated coefficients. *, **, and *** denote significant differences from zero at the 10%, 5%, and 1% levels, respectively.

variable measures because each is a relative measure and hence by construction will not exhibit seasonality. In results not reported, we find the December indicator variable to be positive and significant at the 1% level (estimated coeff. = 0.0036).

Second, we compare the intraquarter variation in the *BHRG* measure for funds showing the highest window dressing (top 10% or 20%) with that of the other funds. Given that *BHRG* is computed using quarter-end holdings, all funds should show a decreasing pattern in the monthly *BHRG* as we move from the first to the third month of a quarter. Further, funds with the highest levels of *BHRG* should show a more pronounced decreasing pattern if they rebalance aggressively with the intention to either window-dress or to pursue a momentum strategy. However, if this intraquarter variation in *BHRG* is driven by window dressing, then it should be greater for funds with the highest *BHRG* in December-end quarters.

We regress the difference between the first and third month *BHRG* on the top 10% (or top 20%) *BHRG* dummy, and on its interaction with a December indicator variable. We report the results under models 1 and 2 in Table 8. For robustness, we also use as the dependent variable the difference between the average monthly *BHRG* of the first two months and that of the third month (see models 3 and 4). If *BHRG* is capturing window dressing and not momentum, we expect the interaction term to be positive. We observe in all four models that the estimated coefficient on the interaction term is positive and significant at the 1% level.

5. Investors' Flow Response to Window Dressing

5.1 Discussion of rationale and empirical predictions

We empirically examine how window dressing can exist in the presence of rational investors. If investors can be misled by funds' WD activity and result in funds rewarded with higher flows, one should expect that all fund managers will engage in such activity. In contrast, if investors are not deceived by window dressing and punish such funds with lower flows, then one would expect that no manager will engage in window dressing. This line of reasoning motivates us to investigate how rational investors respond to window dressing.

As discussed earlier, fund managers may disclose portfolio holdings with a delay up to 60 days following the end of a quarter. At the end of this delay period, investors can observe the reported quarter-end portfolio holdings and attribute any discrepancy with the fund's quarterly performance to either window dressing or to an improvement in their security selection strategy. If a fund manager window-dresses, but then performs well during the delay period, the manager may receive the benefit of doubt from investors who may attribute the good performance to security selection. As a result, this manager will receive higher investor flows between the filing date and the subsequent filing date compared with the managers who do not window-dress. However, if the WD manager fails to achieve good performance during the delay period, the manager is penalized with lower flows compared with non–window dressers. Figure 1 depicts the timeline showing the above sequence of events related to investor flows and performance.[12]

We test this conjecture by collecting data on funds' filing dates from N-Q and N-CSR filings that are electronically available in the SEC's EDGAR database going back to 1994. We compute investor flows between the filing date and the subsequent filing date using a fund's *TNA* on the two dates and return over the period, and scale by *TNA* as of the filing date.[13] We divide these flows by the number of days between the two filing dates to compute average daily flow, which we then multiply by 90 to express on a quarterly basis. We estimate the following regression:

$$\begin{aligned} QtrFlow_{i,t+1} =\; & \gamma_0+\gamma_1 WD\ dummy_{i,t}+\gamma_2 WD\ dummy_{i,t}\times GoodPerf_{i,t} \\ & +\gamma_3 GoodPerf_{i,t}+\gamma_4 QtrAlphaTop_{i,t}+\gamma_5 QtrAlphaMid_{i,t}+\gamma_6 QtrAlphaBot_{i,t} \\ & +\gamma_7 Manager\ Skill_{i,t-1}+\gamma_8 Expense_{i,t}+\gamma_9 Turnover_{i,t}+\gamma_{10} Size_{i,t}+\gamma_{11} Load_{i,t} \\ & +\gamma_{12} Trade\ Cost_{i,t}+Style\ dummies+Time\ dummies+\omega_{i,t} \end{aligned} \tag{11}$$

12 Some funds may decide to voluntarily disclose their portfolio holdings sooner than 60 days. Such a decision is likely to be strategic, as shown by Ge and Zheng (2006). Therefore, to the extent that early voluntary disclosure has information, it should bias us against finding results using the performance during the entire 60-day delay period.

13 When *TNA* is not available for a particular date, we use the closest month-end figure adjusted for the returns between that month end and the filing date.

where $QtrFlow_{i,t+1}$ represents the investor flows for fund i between the fund's filing date in quarter t+1 and its next quarter filing date, $WDdummy_{i,t}$ is an indicator variable that takes a value of 1 when the *Rank Gap* or *BHRG* measure for fund i is either in the top 10% or in the top 20% of all funds in quarter t, $GoodPerf_{i,t}$ is an indicator variable that equals 1 if the average daily return of fund i over the delay period is positive and 0 otherwise, and $QtrAlphaTop_{i,t}$, $QtrAlphaMid_{i,t}$, and $QtrAlphaBot_{i,t}$ are the top 20%, middle 60%, and bottom 20% performance quintiles for fund i in quarter t as defined in Sirri and Tufano (1998). Note that in Equation (11) we use the indicator (rather than the continuous) forms of window dressing and performance so that we can compare the levels of flows between window dressers and non–window dressers.

Our second hypothesis predicts that a manager benefits from window dressing through higher incremental flows between the filing date and the next quarter's filing date if the manager performs well over the delay period. However, the manager incurs a cost in terms of lower incremental flows if performance is poor. Using the coefficient estimates from Equation (11), we illustrate in the following matrix the incremental flows for window dressers and non–window dressers corresponding to good and bad performance during the delay period.

	Window dressers	Non–window dressers	Column diff.
good performance	$\gamma_1+\gamma_2+\gamma_3$	γ_3	$\gamma_1+\gamma_2$
bad performance	γ_1	0	γ_1
Row diff.	$\gamma_2+\gamma_3$	γ_3	

For good performance, the column difference in flows for window dressers and non–window dressers is $(\gamma_1+\gamma_2)$. The first prediction of our second hypothesis is that this difference should be positive. In contrast, if performance is bad, the column difference is simply γ_1. The second prediction is that window dressers will receive lower flows if they perform poorly, thus implying that γ_1 should be negative. Together, these two predictions imply that window dressers are making a risky bet that performance during the delay period will be good. We test this by inspecting the row differences, which provide information on the dispersion in payoffs between the two performance states (good and bad). If window dressers are indeed making a more risky bet, we expect $\gamma_2+\gamma_3>\gamma_3$, or γ_2 to be positive.

5.2 Empirical results

The results from estimating Equation (11), reported in Table 9, are generally consistent with our arguments above. First, the estimated coefficient γ_1 on the *WD Dummy* is negative in the first four specifications, and significant at conventional levels in models 1, 3, and 4 (p-values of 0.001, 0.000, and 0.002, respectively) and close to significant in model 2 (p-value of 0.109). This finding regarding the coefficient γ_1 is consistent with the prediction that window dressers attain lower flows when the delay period performance is poor. Second, the sum of the estimated coefficients on (a) the *WD Dummy* and (b) its interaction with the good performance dummy, $(\gamma_1+\gamma_2)$, is positive in all four specifications and significant in one (see results of the F-test at the bottom of Table 9). This suggests that window dressers obtain higher or no worse flows than non–window dressers if performance is good. Third, estimated coefficient γ_2 is positive and strongly significant in all four specifications consistent with risk-taking by window dressers (i.e., a higher dispersion in flows between good and bad states). This finding, coupled with our earlier results showing that window dressers have lower skill and performance, complements the prior literature that has shown that poorly performing managers have greater career concerns (see Khorana 1996) and exhibit higher risk-taking behavior (see Brown, Harlow, and Starks 1996; Chevalier and Ellison 1997).

5.3 Robustness tests

In this section we conduct three tests to further support the above findings related to our portrayed rationale for window dressing. First, as discussed earlier in Section 2.1.3, there is a potential mechanical relation between the WD measures and past performance. Thus, to show that our WD measures are distinct, we conduct the follow test. We again estimate Equation (11), but instead substitute *AR10%* and *AR20%* (i.e., bottom 10% and 20% performers, respectively) as independent indicator variables to replace the corresponding WD variables. This test is to rule out the alternative hypothesis that investors care nonlinearly about the return patterns in two consecutive quarters (i.e., poor performance during the past quarter and good performance during the delay period) rather than window dressing per se. In results reported in models 5 and 6 of Table 9, we do not find any evidence supporting this alternative hypothesis. Specifically, we find that the estimated coefficient on the interaction between actual returns (*AR10%* or *AR20%*) and the good performance dummy during the delay period is insignificant in both models. This is in sharp contrast to our findings in models 1–4, where the estimated coefficients on the interaction terms are consistently positive and significant. These findings underscore the distinctiveness of our WD measures in explaining how investors resolve the potential conflict between disclosed holdings and quarterly performance.

In our second test, we highlight the specialness of the delay period in our rationale for window dressing. Specifically, in Equation (11), we interact the WD measures with alternative measures of past-quarter performance (instead of the delay-period performance). As reported in Table SA.3 of the Supplementary Appendix, we find that the estimated coefficients on the interaction terms are insignificant in all specifications and often negative. We believe that this is because the interaction of window dressing and past performance does not add to the information set of investors. Rather, the delay period provides investors with new information that helps resolve any conflict between disclosed holdings and quarterly performance.

In our third test, we investigate whether our results are robust after allowing for a change in flow sensitivity to past performance in addition to the flow

Table 9
Window dressing and fund flows

Variables	(1)	(2)	(3)	(4)	(5)	(6)
BHRG10% Dummy$_t$	−0.0244***					
	0.001					
BHRG10% Dummy$_t$ × GoodPerf$_t$	0.0328***					
	0.003					
BHRG20% Dummy$_t$		−0.0098				
		0.109				
BHRG20% Dummy$_t$ × GoodPerf$_t$		0.0269***				
		0.003				
Rank Gap10% Dummy$_t$			−0.0195***			
			0.000			
Rank Gap10% Dummy$_t$ × GoodPerf$_t$			0.0248**			
			0.011			
Rank Gap20% Dummy$_t$				−0.0148***		
				0.002		
Rank Gap20% Dummy$_t$ × GoodPerf$_t$				0.0223***		
				0.010		
AR10% Dummy$_t$					−0.0181***	
					0.002	
AR10% Dummy$_t$ × GoodPerf$_t$					0.0084	
					0.390	
AR20% Dummy$_t$						0.0066
						0.253
AR20% Dummy$_t$ × GoodPerf$_t$						0.0115
						0.203
GoodPerf$_t$	0.0249***	0.0230***	0.0261***	0.0241***	0.0270***	0.0246***
	0.000	0.000	0.000	0.000	0.000	0.000
QtrAlphaBot_t	0.0604*	0.0626*	0.0530	0.0562	0.0242	0.0666*
	0.087	0.077	0.138	0.111	0.492	0.060
QtrAlphaMid_t	0.0345***	0.0358***	0.0339***	0.0334***	0.0337***	0.0309***
	0.000	0.000	0.001	0.001	0.001	0.002
QtrAlphaTop_t	0.0385	0.0389	0.0413	0.0417	0.0428	0.0106
	0.447	0.443	0.415	0.410	0.398	0.841
Manager Skill_t	1.0964	1.0238	1.0549	1.0115	1.0191	0.9984
	0.183	0.213	0.198	0.213	0.216	0.227
Expense_t	0.0620	0.0032	0.0857	0.0675	0.1155	0.0032
	0.939	0.997	0.916	0.934	0.886	0.997
Size_t	−0.0950	−0.0943	−0.1013	−0.0939	−0.0901	−0.1070
	0.473	0.479	0.448	0.481	0.497	0.422
Turnover_t	0.0780**	0.0700**	0.0698**	0.0690**	0.0622*	0.0623*
	0.019	0.039	0.034	0.034	0.055	0.054
Load_t	−0.0172***	−0.0172***	−0.0176***	−0.0175***	−0.0180***	−0.0175***
	0.000	0.000	0.000	0.000	0.000	0.000
Trade Cost_t	−0.0319	−0.0365*	−0.0336	−0.0346	−0.0341	−0.0353
	0.150	0.096	0.127	0.115	0.122	0.109
Constant	0.0147	0.0022	−0.0230	−0.0210	−0.0175	−0.0377
	0.489	0.916	0.300	0.343	0.424	0.100
Time & Style dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,049	3,049	3,054	3,054	3,054	3,054
Adj. R^2	0.0382	0.0372	0.0364	0.0368	0.0361	0.0366
H0: *WD Dummy$_t$* + *WD Dummy$_t$* ×	0.0084	0.0171**	0.0053	0.0075		
GoodPerf$_t$ = 0 p-value	0.40	0.04	0.57	0.32		
H0: *AR Dummy$_t$* + *AR Dummy$_t$* × *GoodPerf$_t$* = 0					−0.0097	0.0181
p–value					0.25	0.02

This table reports the results of regressions where the dependent variable is the fund flows computed on a quarterly basis between the filing dates associated with quarters t and t+1. *GoodPerf* is an indicator variable that takes a value of 1 if a fund's average daily return during the delay period between the quarter-end and filing date is positive, and 0 otherwise. *AR10% (AR20%) Dummy* is an indicator variable that takes a value of 1 if the fund is in the bottom decile (quintile) based on actual returns during a quarter, and 0 otherwise. *QtrAlphaTop$_t$*, *QtrAlphaMid$_t$*, and *QtrAlphaBot$_t$* are the top 20%, middle 60%, and bottom 20% performance quintiles for a fund in quarter t as defined in Sirri and Tufano (1998). Other variables are as defined in Tables 1, 2, and 4. The last row of the table reports the sum of the *WD Dummy* (based on either *Rank Gap* or *BHRG*) or *AR Dummy* and its interaction with the *GoodPerf*, along with the p-values from the F-test for the sum equaling zero. Standard errors are clustered at the fund level. p-values are reported below the estimated coefficients. *, **, and *** denote significant differences from zero at the 10%, 5%, and 1% levels, respectively.

sensitivity to the delay-period performance. For this purpose, we modify Equation (11) and include a three-way interaction term of *WD Dummy*, delay period performance, and past performance. In results reported in Table SA.4 of the Supplementary Appendix, the estimated coefficient on the interaction term is positive but insignificant in three of the four specifications, while insignificantly negative in the other. This finding suggests that there is little evidence of an increase in flow sensitivity to past performance due to window dressing. More importantly, the estimated coefficient on the two-way interaction between *WD Dummy* and delay-period performance continues to be positive and strongly significant in all four specifications. This shows that our main finding of investors' incremental flow response to window dressing and delay-period performance continues to hold.

Taken together, the results are consistent with our rationale for window dressing. If fund performance during the delay period is good, investors attribute the incongruence between holdings-based returns and actual returns to improved security selection and reward the fund with higher flows. In contrast, if fund performance is bad, then investors attribute the incongruence to window dressing and punish the fund with lower flows.

6. Window Dressing and Future Performance

Our above findings indicate that WD managers are generally value-destroyers due to their excessive portfolio turnover and trade cost generation. Further, such managers appear to be unskilled and trade for noninformational reasons. As such we hypothesize that WD managers should exhibit, on average, poor future performance. To test this hypothesis, we analyze the future performance of window dressers over the short term (i.e., the subsequent quarter) as well as over the long term (i.e., one, two, and three years) after controlling for past performance (including poor past performance).

6.1 Future short-term performance

6.1.1 Sorts analyses. We begin with a univariate analysis where funds are sorted each quarter into WD deciles. Using daily returns and trade information (as imputed from changes in holdings), we estimate funds' quarterly alphas and total trade costs over the next quarter. For each decile, we compute and report means of these variables in Table 10 (see Panel A for *BHRG* and Panel B for *Rank Gap*). For each WD measure, we observe that the alphas exhibit a generally monotonically *decreasing* pattern as we move from the lowest to the highest decile. This pattern suggests that greater window dressing is associated with lower subsequent performance. For the highest WD decile, next quarter's average alphas are −0.66% and −0.74% for *BHRG* and *Rank Gap*, respectively. Further, the differences in the mean alphas between the bottom and top deciles (10–1) are −0.63% and −0.57% for *BHRG* and *Rank Gap*, respectively, and both are highly significant. We also observe that window dressers continue

Table 10
Window dressing and next-quarter fund performance: Univariate results

Panel A

	Decile	*3-Month Alpha*	p-value	*Trade Cost*	p-value	*Actual Return*	p-value	*Momentum Beta*	p-value
	1 (low)	−0.0003	0.70	0.0016	0.00	0.0066	0.00	−0.0287	0.00
	2	−0.0025	0.00	0.0013	0.00	0.0056	0.00	−0.0141	0.00
	3	−0.0031	0.00	0.0012	0.00	0.0101	0.00	−0.0103	0.00
BHRG	4	−0.0037	0.00	0.0011	0.00	0.0151	0.00	−0.0041	0.16
	5	−0.0030	0.00	0.0010	0.00	0.0131	0.00	−0.0040	0.13
	6	−0.0025	0.00	0.0011	0.00	0.0120	0.00	0.0097	0.00
	7	−0.0037	0.00	0.0014	0.00	0.0074	0.00	0.0316	0.00
	8	−0.0035	0.00	0.0016	0.00	0.0057	0.00	0.0576	0.00
	9	−0.0047	0.00	0.0021	0.00	0.0013	0.35	0.0909	0.00
	10 (high)	−0.0066	0.00	0.0031	0.00	−0.0013	0.41	0.1781	0.00
	10–1	−0.0063	0.00	0.0014	0.00	−0.0078	0.00	0.2068	0.00

Panel B

	Decile	*3-Month Alpha*	p-value	*Trade Cost*	p-value	*Actual Return*	p-value	*Momentum Beta*	p-value
	1 (low)	−0.0017	0.01	0.0015	0.00	0.0122	0.00	0.0139	0.00
	2	−0.0022	0.00	0.0014	0.00	0.0129	0.00	0.0272	0.00
	3	−0.0034	0.00	0.0013	0.00	0.0103	0.00	0.0300	0.00
Rank Gap	4	−0.0029	0.00	0.0013	0.00	0.0081	0.00	0.0268	0.00
	5	−0.0030	0.00	0.0013	0.00	0.0071	0.00	0.0207	0.00
	6	−0.0040	0.00	0.0014	0.00	0.0065	0.00	0.0168	0.00
	7	−0.0032	0.00	0.0014	0.00	0.0034	0.01	0.0189	0.00
	8	−0.0026	0.00	0.0015	0.00	0.0051	0.00	0.0221	0.00
	9	−0.0030	0.00	0.0018	0.00	0.0041	0.00	0.0339	0.00
	10 (high)	−0.0074	0.00	0.0024	0.00	0.0030	0.04	0.0971	0.00
	10–1	−0.0057	0.00	0.0009	0.00	−0.0093	0.00	0.0832	0.00

This table reports the means of next-quarter *3-Month Alpha*, *Trade Cost*, *Actual Return*, and current-quarter *Momentum Beta* for decile portfolios of mutual funds sorted by the two WD measures (*BHRG* in Panel A and *Rank Gap* in Panel B) over a quarter. Current-quarter *Momentum Beta* is the average of the three monthly momentum betas. The monthly momentum beta is the estimated coefficient on the momentum factor obtained from the regression of daily net-of-fee fund returns on the returns of the four factors (excess market returns, size and book-to-market factors, and the momentum factor) for the month. Other variables are as defined in Table 1. p-values of the t-tests are reported after adjusting the standard errors for clustering at the fund level.

to generate high levels of trade costs in the subsequent quarter, which appear to contribute significantly to their lower performance. In decile 10 of both panels, next quarter trade costs are 0.31% and 0.24% for *BHRG* and *Rank Gap*, respectively. Also, the (10–1) spreads in trade cost are large and significant.

Since buying winners and selling losers can also be attributed to momentum trading, we also report the next quarter's mean returns (*AR*) and the current quarter's momentum betas. Interestingly, in both panels the momentum betas show a monotonically *increasing* pattern, ranging from −0.0287 (0.0139) for the lowest decile to 0.1781 (0.0971) for the highest decile of funds sorted by *BHRG* (*Rank Gap*). This increasing pattern of the momentum betas should be, on average, associated with an increasing pattern of fund returns rather than the generally decreasing pattern that we observe. Specifically, the returns in both panels are lowest for the highest WD deciles with the (10−1) spreads being statistically significant at −0.78% and −0.93% for *BHRG* and *Rank Gap*, respectively. These results further support our contention that window dressers, though appearing to follow a momentum-like strategy, are actually destroying value with their portfolio churning and information-less trading.

For robustness, we examine whether the above finding that window dressing is associated with poor future performance continues to hold after controlling for past performance, as there could be persistence in fund performance, especially when prior performance has been poor (e.g., see Brown and Goetzmann 1995; Carhart 1997). To investigate this, we double-sort funds on window dressing and past performance into 5 × 5 quintiles. In results reported in Table SA.5 of the Supplementary Appendix, we find that (5–1) spread in future performance is the most negative for funds in the highest WD quintile and lowest past performance. Thus, controlling for past performance does not change our finding that window dressing is negatively related to future performance.

To further shed light on the source of the poor future performance of window dressers, we double-sort funds on window dressing and next-quarter trade cost, and examine next-quarter performance. We report these findings in Table 11. In both panels, the worst performance is associated with funds in the highest WD and highest trade-cost quintile (*3-Month Alpha* equal to −0.0109 in Panel A, and −0.0111 in Panel B). Further, the (5–1) spread in future performance is the most negative for funds in the highest WD quintile (−0.0079 in Panel A and −0.0102 in Panel B). In contrast, for funds in the lowest WD quintile and highest trade-cost quintile, future performance is the best (*3-Month Alpha* equal to −0.0006 in Panel A and +0.0016 in Panel B). Also, the (5–1) spread in future performance for funds in the lowest WD quintile is positive (0.0004 in Panel A and 0.0037 in Panel B), but only statistically significant in Panel B. These results indicate that the poor future performance of window dressers is partly driven by their high trade costs.[14]

6.1.2 Regression analysis. We further examine the relation between window dressing and future performance in a multivariate setting where we control for manager skill, past performance (including poor past performance), and various fund attributes. We estimate the following regression:

$$\begin{aligned}\alpha_{i,t+1} = {} & \vartheta_0 + \vartheta_1 WD_{i,t} + \vartheta_2 Perf_{i,t} + \vartheta_3 AR_{i,t} + \vartheta_4 Manager\ Skill_{i,t-1} \\ & + \vartheta_5 Expense_{i,t+1} + \vartheta_6 Turnover_{i,t+1} + \vartheta_7 Size_{i,t+1} + \vartheta_8 Load_{i,t+1} + \vartheta_9 Flow_{i,t+1} \\ & + \vartheta_{10} Trade\ Cost_{i,t+1} + Style\ dummies + Time\ dummies + \varepsilon_{i,t+1} \qquad (12)\end{aligned}$$

where $\alpha_{i,t+1}$ is fund i's quarterly alpha during quarter $t+1$, $WD_{i,t}$ is fund i's WD measure in quarter t, $Perf_{i,t}$ represents the past performance (either one-month alpha, three-month alpha, one-year alpha, or performance rank) of fund i during quarter t, $AR_{i,t}$ is an indicator variable that takes a value of 1 if the actual returns

[14] We also repeat this analysis by using current-quarter trade cost instead of next-quarter trade cost to investigate whether it predicts next-quarter performance. Results reported in Table SA.6 of the Supplementary Appendix show that for window dressers, higher trading costs in the current quarter are associated with lower future performance.

Table 11
Next-quarter fund performance: Results from double 5 × 5 sorts on window dressing and next-quarter trade costs

Panel A: *BHRG*

		Next-quarter *Trade Cost*					
		1 (low)	2	3	4	5 (high)	5–1
	1 (low)	−0.0010	−0.0028	−0.0026	−0.0010	−0.0006	0.0004
		0.29	0.00	0.00	0.28	0.60	0.79
	2	−0.0029	−0.0036	−0.0033	−0.0041	−0.0028	0.0001
		0.00	0.00	0.00	0.00	0.00	0.95
BHRG	3	−0.0007	−0.0022	−0.0025	−0.0029	−0.0048	−0.0040
		0.20	0.00	0.00	0.00	0.00	0.00
	4	−0.0015	−0.0029	−0.0042	−0.0047	−0.0033	−0.0018
		0.03	0.00	0.00	0.00	0.00	0.12
	5 (high)	−0.0031	−0.0041	−0.0057	−0.0097	−0.0109	−0.0079
		0.00	0.00	0.00	0.00	0.00	0.00
	5–1	−0.0021	−0.0012	−0.0031	−0.0087	−0.0103	
		0.11	0.28	0.01	0.00	0.00	

Panel B: *Rank Gap*

		1 (low)	2	3	4	5 (high)	5–1
	1 (low)	−0.0021	−0.0038	−0.0035	−0.0021	0.0016	0.0037
		0.32	0.00	0.00	0.02	0.15	0.01
	2	−0.0023	−0.0040	−0.0029	−0.0039	−0.0027	−0.0004
		0.00	0.00	0.00	0.00	0.01	0.78
Rank Gap	3	−0.0029	−0.0036	−0.0034	−0.0050	−0.0045	−0.0016
		0.00	0.00	0.00	0.00	0.00	0.21
	4	−0.0010	−0.0015	−0.0040	−0.0050	−0.0049	−0.0040
		0.20	0.20	0.00	0.00	0.00	0.00
	5 (high)	−0.0009	−0.0028	−0.0047	−0.0066	−0.0111	−0.0102
		0.23	0.00	0.00	0.00	0.00	0.00
	5–1	0.0012	0.0009	−0.0012	−0.0044	−0.0127	
		0.31	0.38	0.26	0.00	0.00	

This table reports means of next quarter *3-Month Alpha* for twenty-five portfolios of mutual funds sorted first by the two WD measures (*BHRG* in Panel A and *Rank Gap* in Panel B) and then by *Trade Cost* during the next quarter. All variables are as defined in Table 1. p-values of the t-tests are reported below the means after adjusting the standard errors for clustering at the fund level.

of fund i during quarter t are in the bottom decile and 0 otherwise, and all other variables are as defined previously. We report the results from this estimation in Table 12 for WD measures *BHRG10%* and *Rank Gap10%*, respectively.[15]

We find that the estimated coefficients on *BHRG10%* are negative and highly significant in models 1−4. Further, based on the magnitudes of the estimated coefficients, window dressers' performance in the subsequent quarter appears to range from 40−60 basis points lower on average. In models 5−8, the results for the estimated coefficients on *Rank Gap10%* are negative in all four specifications (ranging from 20−30 basis points) and are significant in three out of the four cases. Unlike window dressing, we find positive and significant coefficients on the *AR10%* in each model. This suggests that window dressing is distinct from poor past performance in explaining future performance. Also, consistent with KSZ, we observe that manager skill is positively related to future

15 Given that we have six alternative WD measures and four measures of past performance, for sake of brevity we present in Table 12 only the results for *BHRG10%* and *Rank Gap10%*. The findings for the other WD measures are qualitatively similar.

Table 12
Window dressing and next-quarter fund performance: Multivariate results

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
BHRG10% Dummy$_t$	−0.0052∗∗∗	−0.0043∗∗∗	−0.0058∗∗∗	−0.0056∗∗∗				
	0.000	0.000	0.000	0.000				
Rank Gap10% Dummy$_t$					−0.0028∗∗∗	−0.0014	−0.0018∗	−0.0029∗∗∗
					0.002	0.115	0.052	0.002
AR10% Dummy$_t$	0.0022∗∗	0.0046∗∗∗	0.0031∗∗∗	0.0027∗∗	0.0029∗∗∗	0.0050∗∗∗	0.0037∗∗∗	0.0025∗∗
	0.027	0.000	0.002	0.011	0.004	0.000	0.000	0.018
Manager Skill$_{t\text{-}1}$	0.2743∗∗	0.2343∗∗		0.2792∗∗	0.2885∗∗∗	0.2440∗∗		0.2916∗∗∗
	0.012	0.027		0.012	0.009	0.024		0.008
Expense$_{t+1}$	−0.0031	−0.0204	−0.0726	−0.0097	−0.0340	−0.0531	−0.1066	−0.0313
	0.968	0.778	0.309	0.901	0.663	0.476	0.147	0.690
Size$_{t+1}$	0.0317∗∗	0.0311∗∗	0.0059	0.0246∗	0.0202	0.0209	−0.0024	0.0204
	0.031	0.028	0.673	0.097	0.171	0.145	0.862	0.170
Turnover$_{t+1}$	−0.0022	−0.0018	0.0029	−0.0027	−0.0038	−0.0033	0.0003	−0.0038
	0.536	0.610	0.399	0.439	0.286	0.344	0.933	0.279
Load$_{t+1}$	−0.0007	−0.0006	−0.0003	−0.0007	−0.0005	−0.0004	−0.0002	−0.0006
	0.185	0.242	0.508	0.201	0.330	0.407	0.772	0.311
Flow$_{t+1}$	0.0460∗∗∗	0.0435∗∗∗	0.0425∗∗∗	0.0448∗∗∗	0.0449∗∗∗	0.0425∗∗∗	0.0412∗∗∗	0.0452∗∗∗
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Trade Cost$_{t+1}$	−0.0023	−0.0033	−0.0059∗∗	−0.0016	−0.0033	−0.0044∗	−0.0066∗∗∗	−0.0032
	0.336	0.167	0.015	0.514	0.172	0.073	0.007	0.183
1-Month Alpha$_t$	0.0360∗∗				0.0257∗			
	0.014				0.093			
3-Month Alpha$_t$		0.0811∗∗∗				0.0766∗∗∗		
		0.000				0.000		
1-Year Alpha$_t$			0.0403∗∗∗				0.0392∗∗∗	
			0.000				0.000	
PerfRank$_t$				−0.0000				0.0000
				0.437				0.880
Constant	−0.0024	−0.0026	−0.0011	−0.0014	−0.0014	−0.0017	−0.0004	−0.0015
	0.185	0.130	0.532	0.464	0.449	0.344	0.813	0.439
Time & Style dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30,512	29,777	28,069	29,356	29,408	28,672	27,155	29,408
Adj. R^2	0.0597	0.0656	0.0721	0.0593	0.0585	0.0641	0.0696	0.0583

This table reports the results of regressions of the *3-Month Alpha* on the two WD measures (*BHRG10% Dummy* and *Rank Gap10% Dummy*), *AR10% Dummy*, and various fund attributes. The dependent variable is *3-Month Alpha* during the leading quarter. The *1-Month Alpha* and *1-Year Alpha* are defined as the sum of four-factor daily alphas over the quarter-end month, and over a year up to the quarter-end, respectively. *PerfRank* is the percentile rank of a fund's quarterly performance. Other variables are as defined in Tables 1, 4, and 9. Standard errors are adjusted for clustering at the fund level. p-values are reported below the estimated coefficients. *, **, and *** denote significant differences from zero at the 10%, 5%, and 1% levels, respectively.

performance. Together, the findings in this section support our third hypothesis that window dressing is associated with lower future performance.

6.2 Long-term performance

We sort funds each quarter into WD deciles, and compute the average one-, two-, and three-year alphas for each fund portfolio. The results, presented in Table 13, indicate a notable drop in fund performance in the highest WD deciles. To illustrate, for the *BHRG*-based results presented in Panel A, the one-, two-, and three-year alphas in decile 10 are −3.04%, −6.91%, and −11.36%, respectively. Further, the (10–1) decile differences are also all negative and highly significant. Similar findings are observed in Panel B based on the *Rank Gap* measure.

To show that the above findings are distinct from poor past performance, we also sort funds on actual return (*AR*) and similarly compute the average alphas over the next one, two, and three years. The results are reported in Panel C of Table 13.[16] We find that the spreads in the future long-term alphas are also negative and significant; however, the magnitude of the spreads is less negative than those reported in Panels A and B. We test this by conducting a difference-in-differences test based on (a) the spread between the extreme deciles of each WD measure, and (b) the spread between extreme *AR* deciles. As reported in Panel D, the difference-in-differences are significantly negative, indicating that the spreads between the extreme deciles of window dressing are more negative.[17] We also analyze this issue in a multivariate setting by estimating the regression specification given in Equation (12) above, but using as alternative dependent variables funds' one-, two-, and three-year alphas. In results reported in Table SA.8 of the Supplementary Appendix, we continue to find that window dressing dominates poor past performance in explaining future long-term performance.

7. Concluding Remarks

We shed light on alleged window-dressing behavior of mutual fund managers using a large sample of U.S. equity funds over the period September 1998 to December 2008. We develop relative and absolute measures of window dressing to capture the inconsistency between the information based on realized fund returns and that based on disclosed portfolio holdings. We conduct a battery of tests to show that these measures capture window-dressing behavior rather than momentum trading. Providing further support

[16] Since window dressing is associated with poor past performance, we reverse the sorting order in Panel C such that the lowest (highest) *AR* decile in Panel C corresponds to the highest (lowest) WD decile in Panels A and B.

[17] We also conduct a double-sort analysis where funds are sorted by window dressing and past performance (*AR*) into 5x5 quintiles. As reported in Table SA.7 of the Supplementary Appendix, we continue to find that window dressing is negatively related to future long-term performance.

Table 13
Window dressing and long-term fund performance: Univariate results

Panel A: *BHRG*

BHRG decile	*1-Year Alpha*	*2-Year Alpha*	*3-Year Alpha*
1 (low)	−0.0094	−0.0189	−0.0325
2	−0.0116	−0.0221	−0.0353
3	−0.0103	−0.0207	−0.0326
4	−0.0121	−0.0233	−0.0358
5	−0.0115	−0.0231	−0.0341
6	−0.0115	−0.0254	−0.0372
7	−0.0131	−0.0258	−0.0377
8	−0.0162	−0.0327	−0.0492
9	−0.0194	−0.0401	−0.0607
10 (high)	−0.0304	−0.0691	−0.1136
10–1	−0.0210	−0.0502	−0.0811
p-value	0.00	0.00	0.00

Panel B: *Rank Gap*

Rank Gap decile	*1-Year Alpha*	*2-Year Alpha*	*3-Year Alpha*
1 (low)	−0.0069	−0.0123	−0.0286
2	−0.0073	−0.0187	−0.0343
3	−0.0100	−0.0222	−0.0373
4	−0.0112	−0.0237	−0.0393
5	−0.0126	−0.0274	−0.0421
6	−0.0158	−0.0309	−0.0478
7	−0.0154	−0.0332	−0.0530
8	−0.0160	−0.0369	−0.0560
9	−0.0176	−0.0396	−0.0612
10 (high)	−0.0332	−0.0601	−0.0873
10–1	−0.0263	−0.0477	−0.0587
p-value	0.00	0.00	0.00

Panel C: *AR*

AR decile	*1-Year Alpha*	*2-Year Alpha*	*3-Year Alpha*
1 (high)	−0.0076	−0.0143	−0.0323
2	−0.0089	−0.0221	−0.0392
3	−0.0089	−0.0234	−0.0412
4	−0.0125	−0.0264	−0.0440
5	−0.0121	−0.0280	−0.0473
6	−0.0141	−0.0287	−0.0473
7	−0.0149	−0.0311	−0.0487
8	−0.0183	−0.0358	−0.0541
9	−0.0221	−0.0442	−0.0612
10 (low)	−0.0266	−0.0512	−0.0675
10–1	−0.0190	−0.0369	−0.0352
p-value	0.00	0.00	0.00

Panel D: Difference-in-differences (10–1)

	1-Year Alpha	*2-Year Alpha*	*3-Year Alpha*
BHRG – AR	−0.0020	−0.0133	−0.0459
p-value	0.59	0.03	0.00
Rank Gap – AR	−0.0073	−0.0108	−0.0235
p-value	0.00	0.01	0.00

This table reports the means of long-term mutual fund performance subsequent to their WD activity for decile portfolios of mutual funds sorted by the two WD measures (*BHRG* in Panel A and *Rank Gap* in Panel B), and actual returns (*AR* in Panel C) for each quarter. The *1-Year Alpha* is defined as the sum of four-factor daily alphas over the one year up to the quarter-end. The *2-Year Alpha* is defined as the sum of four-factor daily alphas over the two years up to the quarter-end. The *3-Year Alpha* is defined as the sum of four-factor daily alphas over the three years up to the quarter-end. *p*-values of the *t*-tests of the differences for deciles 10 and 1 are reported in the last two rows of Panels A to C after adjusting the standard errors for clustering at the fund level. *p*-values of the *t*-tests of the difference-in-differences tests (between the 10–1 spreads in alphas for *BHRG* [or *Rank Gap*] versus those for *AR*) are reported in Panel D after adjusting the standard errors for clustering at the fund level.

to these measures, we show that window dressing is associated with managers who are less skilled and who perform poorly. Further, we find that window dressing is value-destroying and is associated, on average, with lower future performance. Our study also contributes to the debate on mandatory portfolio disclosure by institutional investors by highlighting some of the unintended consequences.

Given these negative aspects of window dressing, we offer and test a rationale for why managers may engage in such activity. Specifically, we show how window-dressing managers potentially benefit from the delay allowed to disclose portfolio holdings. We argue that if such managers' performance is good during the delay period, investors may attribute the disclosed holdings, tilted toward winner stocks and away from loser stocks, to security selection ability and reward the managers with higher flows. In contrast, if the managers' performance during the delay period turns out to be poor, investors may attribute the disclosed holdings to window dressing and withdraw their capital. These costs and benefits of window dressing show how it can exist in the presence of rational investors who respond to signals of managerial ability inferred from the fund's performance and from disclosed portfolio holdings. Our empirical findings support such an explanation.

Supplementary Data

Supplementary materials for this article are available online at http://rfs.oxfordjournals.org/.

References

Agarwal, V., N. D. Daniel, and N. Y. Naik. 2011. Do hedge funds manage their reported returns? *Review of Financial Studies* 24:3281–3320.

Agarwal, V., W. Jiang, Y. Tang, and B. Yang. 2013. Uncovering hedge fund skill from the portfolios they hide. *Journal of Finance* 68:739–83.

Agarwal, V., K. Mullally, Y. Tang, and B. Yang. Forthcoming. Mandatory portfolio disclosure, stock liquidity, and mutual fund performance. *Journal of Finance*.

Alexander, G. J., G. Cici, and S. Gibson. 2007. Does motivation matter when assessing trade performance? An analysis of mutual funds. *Review of Financial Studies* 20:125–50.

Aragon, G. O., M. Hertzel, and Z. Shi. 2013. Why do hedge funds avoid disclosure? Evidence from confidential 13F filings. *Journal of Financial and Quantitative Analysis* 48:1499–1518.

Baker, M., L. Litov, J. A. Wachter, and J. Wurgler. 2010. Can mutual fund managers pick stocks? Evidence from their trades prior to earnings announcements. *Journal of Financial and Quantitative Analysis* 45:1111–31.

Brown, K. C., W. V. Harlow, and L. T. Starks. 1996. Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry. *Journal of Finance* 51:85–110.

Brown, S. J., and W. N. Goetzmann. 1995. Performance persistence. *Journal of Finance* 50:679–98.

Brown, S. J., and C. Schwarz. 2011. The impact of mandatory hedge fund portfolio disclosure. Working Paper, New York University and University of California at Irvine.

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

Carhart, M. M., R. Kaniel, D. K. Musto, and A. V. Reed. 2002. Leaning for the tape: Evidence of gaming behavior in equity mutual funds. *Journal of Finance* 58:661–93.

Chen, H., N. Jegadeesh, and R. Wermers. 2000. The value of active mutual fund management: An examination of the stockholdings and trades of fund managers. *Journal of Financial and Quantitative Analysis* 35:343–68.

Chen, H., and V. Singal. 2004. All things considered, taxes drive the January effect. *Journal of Financial Research* 27:351–72.

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

Cohen, R. B., J. D. Coval, and L. Pastor. 2005. Judging fund managers by the company they keep. *Journal of Finance* 60:1057–94.

Cremers, K. J. M., and A. Petajisto. 2009. How active is your fund manager? A new measure that predicts performance. *Review of Financial Studies* 22:3329–65.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance* 52:1035–58.

Frank, M. M., J. M. Poterba, D. A. Shackelford, and J. B. Shoven. 2004. Copycat funds: Information disclosure regulation and the returns to active management in the mutual fund industry. *Journal of Law and Economics* 47:515–41.

Ge, W., and L. Zheng. 2006. The frequency of mutual fund portfolio disclosure. Working Paper, University of Washington and University of California at Irvine.

Gompers, P. A., and A. Metrick. 2001. Institutional investors and equity prices. *Quarterly Journal of Economics* 116:229–59.

Grinblatt, M., and S. Titman. 1989. Mutual fund performance: An analysis of quarterly portfolio holdings. *Journal of Business* 62:394–416.

———. 1993. Performance measurement without benchmarks: An examination of mutual fund returns. *Journal of Business* 66:47–68.

Grinblatt, M., S. Titman, and R. Wermers. 1995. Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior. *American Economic Review* 85:1088–1105.

Haugen, R. A., and J. Lakonishok. 1988. The incredible January effect: The stock market's unsolved mystery. Homewood, Illinois: Dow Jones-Irwin.

He, J., L. Ng, and Q. Wang. 2004. Quarterly trading patterns of financial institutions. *Journal of Business* 77:493–509.

Hu, G., R. D. McLean, J. Pontiff, and Q. Wang. 2014. The year-end trading activities of institutional investors: Evidence from daily trades. *Review of Financial Studies* 27:1593–1614.

Hu, P., J. R. Kale, M. Pagani, and A. Subramanian. 2011. Fund flows, performance, managerial career concerns, and risk-taking. *Management Science* 57:628–46.

Huang, J., K. D. Wei, and H. Yan. 2007. Participation costs and the sensitivity of fund flows to past performance. *Journal of Finance* 62:1273–1311.

Huang, L., and J. R. Kale. 2013. Product market linkages, manager quality, and mutual fund performance. *Review of Finance* 17:1895–1946.

Ippolito, R. A. 1992. Consumer reaction to measures of poor quality: Evidence from the mutual fund industry. *Journal of Law and Economics* 35:45–70.

Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48:65–91.

Jiang, G. J., T. Yao, and T. Yu. 2007. Do mutual funds time the market? Evidence from portfolio holdings. *Journal of Financial Economics* 86:724–58.

Kacperczyk, M., and A. Seru. 2007. Fund manager use of public information: New evidence on managerial skills. *Journal of Finance* 62:485–528.

Kacperczyk, M., C. Sialm, and L. Zheng. 2005. On the industry concentration of actively managed equity funds. *Journal of Finance* 60:1983–2012.

———. 2008. Unobserved actions of mutual funds. *Review of Financial Studies* 21:2379–2416.

Khorana, A. 1996. Top management turnover: An empirical investigation of mutual fund managers. *Journal of Financial Economics* 40:403–27.

Lakonishok, J., A. Shleifer, R. Thaler, and R. Vishny. 1991. Window dressing by pension fund managers. *American Economic Review* 81:227–31.

Meier, I., and E. Schaumburg. 2004. Do funds window dress? Evidence for U.S. domestic equity mutual funds. Working Paper, HEC Montreal and Kellogg School of Management.

Morey, M. R., and E. O'Neal. 2006. Window dressing in bond mutual funds. *Journal of Financial Research* 29:325–47.

Moskowitz, T. J. 2000. Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses: Discussion. *Journal of Finance* 55:1695–1703.

Musto, D. K. 1997. Portfolio disclosure and year-end price shift. *Journal of Finance* 52:1563–88.

———. 1999. Investment decisions depend on portfolio disclosures. *Journal of Finance* 54:935–52.

Ng, L., and Q. Wang. 2004. Institutional trading and the turn-of-the-year effect. *Journal of Financial Economics* 74:343–66.

Poterba, J. M., and S. J. Weisbenner. 2001. Capital gains tax rules, tax-loss trading, and turn-of-the-year returns. *Journal of Finance* 56:353–68.

Ritter, J. R., and N. Chopra. 1989. Portfolio rebalancing and the turn-of-the-year effect. *Journal of Finance* 44:149–66.

Sias, R. W. 2007. Causes and seasonality of momentum profits. *Financial Analysts Journal* 63:48–54.

Sias, R. W., and L. T. Starks. 1997. Institutions and individuals at the turn-of-the-year. *Journal of Finance* 52:1543–62.

Sias, R. W., L. T. Starks, and S. Titman. 2006. Changes in institutional ownership and stock returns: Assessment and methodology. *Journal of Business* 79:2869–2910.

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

Solomon, D. H., E. F. Soltes, and D. Sosyura. 2014. Winners in the spotlight: Media coverage of fund holdings as a driver of flows. *Journal of Financial Economics* 113:53–72.

Starks, L. T., L. Yong, and L. Zheng. 2006. Tax-loss selling and the January effect: Evidence from municipal bond closed-end funds. *Journal of Finance* 61:3049–67.

Verbeek, M., and Y. Wang. 2013. Better than the original? The relative success of copycat funds. *Journal of Banking & Finance* 37:3454–71.

Wermers, R. 1999. Mutual fund herding and the impact on stock prices. *Journal of Finance* 54:581–622.

———. 2000. Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses. *Journal of Finance* 55:1655–95.

———. 2001. The potential effects of more frequent portfolio disclosure on mutual performance. *Investment Company Institute Perspective* 7:1–12.