# Do Funds Window Dress?

# Evidence for U.S. Domestic Equity Mutual Funds*

Iwan Meier[a] and Ernst Schaumburg[b]

[a]HEC Montréal

[b]Kellogg School of Management

May 1, 2006

**Abstract**

It is not uncommon for mutual fund managers to make significant adjustments to their allocations before closing out their portfolios at the end of the quarter. The common wisdom on the street and in the financial press is that part of this activity is due to "window dressing" which has become widespread as investors are becoming increasingly sophisticated in analyzing fund holdings as well as past returns in an effort to detect manager skill. So far, the academic literature on equity mutual funds has provided little empirical evidence to this effect. We analyze the semi-annual holdings and daily net asset values of 4,025 U.S. domestic equity mutual funds over the period from 1997 to 2002 and find strong evidence of increased turnover during the last days of the quarter, consistent with window dressing. In particular, we show that growth funds and funds with poor recent performance are more likely to report misleading holding. Furthermore, the end of quarter trading activity is not easily accounted for by momentum/relative strength strategies and is not associated with strategies that on average provide any added value to investors, even before accounting for expenses. Nor can liquidity costs explain these findings.

*We thank Annette Larson, Paul Kaplan, Robert Korajczyk, Robert McDonald, Todd Pulvino, Paula Tkac, participants at the Zell Center for Risk Research seminar at the Kellogg School of Management, the 2004 Hawaii International Conference on Business, the 2004 European Financial Management Association meetings, the JFE Conference on Delegated Portfolio Management in Eugene (Oregon), and seminar participants at HEC Montréal, and Tilburg University for their comments and suggestions. We would especially like to thank Ravi Jagannathan for numerous suggestions and many insightful discussions. We thank Morningstar, Inc. for providing us with the mutual fund data.

# 1 Introduction

The fundamental problem of a mutual fund investor is detecting whether a fund manager has skill. At his disposal the investor has time series data on daily fund returns as well as intermittent reportings of portfolio holdings. The U.S. Securities and Exchange Commission (SEC) requires funds to report their portfolio holdings to shareholders at least semi-annually but most funds voluntarily disclose their portfolio holdings every quarter.[1] These reported holdings are snapshots of the fund portfolio at a given date and may only provide limited information about the funds' actual portfolio positions throughout the quarter. A few funds do release limited information on their top holdings on a more frequent basis, although fund managers never reveal when assets were acquired or sold.[2]

Given the volatility of daily portfolio returns, it is extremely hard to statistically identify superior ability based on past returns alone. Mutual fund rating companies and sophisticated investors increasingly scrutinize reported holdings for evidence of stock selection and/or market timing ability. This may tempt a fund manager to adjust his portfolio immediately before disclosing to the public, e.g. in order to remove particularly embarrasing positions from his book or hide his successful bets from the competition. This deceptive practice is called "window dressing" and leads to reported holdings that are not representative of the actual portfolio held during the quarter. The financial press often attributes end-of-quarter stock price movements and increases in trading volume to window dressing activity by equity mutual funds.[3] While the academic literature has long recognized its existence, e.g. Haugen and Lakonishok (1988) or Lakonishok, Shleifer, Thaler and Vishny (1991), there has been little empirical evidence provided thus far. Lakonishok, Shleifer, Thaler and Vishny (1991) analyze pension fund managers. The authors provide evidence that pension plans dump stocks deemed to be "mistakes" at the end of each quarter, although the practice is most pronounced in December and for small funds. By contrast, the window dressing activity documented in this paper does not appear to be tax driven

[1] The Rules adopted under the Investment Company Act of 1940 requires funds to transmit a record of the holdings to shareholders at least semi-annually [Rule 30e-1(a)], and within 60 days after the end of the reporting period [Rule 30e-1(c)]. The Securities Act of 1933 requires that funds provide a prospectus. For details see University of Cincinnati, Center for Corporate Law, "Securities Lawyer's Deskbook": www.law.uc.edu/CCL/.

[2] Fidelity e.g. updates the top 10 holdings for all of its funds quarterly and the sector weightings and asset allocation monthly. Vanguard provides both information on a monthly basis. Few funds even went one step further: The San Francisco based Metamarkets.com with its open-disclosure fund OpenFund reported all transactions after completion. Metamarkets closed in July 2001, two years after inception.

[3] An example is the following extract from the *Wall Street Journal*, "Dow Advances 67.16; Index Shows Gains As the Quarter Ends", September 30, 2003: "[...] Traders attributed a good part of the day's stock gain to 'window dressing' – an effort by money managers to make their portfolios look more attractive before closing them out for the quarter. Managers with too much uninvested cash feel the need to invest it now. They were buying some of the quarter's biggest winners [...]."

and is not limited to any particular quarter. Musto (1999) performs a test of window dressing by money market funds. From the analysis of weekly portfolio statistics he concludes that money market fund managers hold fewer private issues and more low-risk government bonds around the reporting date. Sias and Starks (1997) examine the trading activity of individual and institutional investors at year-end and find that institutions are more inclined to buy recent winners, which is consistent with the window dressing hypothesis.[4]

In this paper we observe a significant number of fund filings that are preceeded by abnormal turnover during the days leading up to reporting. We propose a methodology to identify reported holdings which are likely not representative based on a comparison of the realized fund return with the hypothetical return of a buy-and-hold strategy. This is achieved by calculating the return the fund would have earned had it held the reported portfolio during the weeks leading up to the reporting date. Contrasting this hypothetical return with the realized, pre-expense net asset value (NAV) based return reveals information about whether a fund manager adjusted his positions in a significant way prior to disclosure.

Based on this identification strategy, we examine a sample of 4,025 U.S. domestic equity mutual funds with a total of 27,702 filings over the period 1997-2002. To identify filings which may not be representative, we require that at least 95% of the portfolio holdings at each filing date can be matched with daily return data from CRSP. This reduces the sample to 3,289 funds with a total of 18,139 filings and a combined market capitalization of $2.0 trillion as of December 2001. This represents 54.5% of the U.S. domestic equity fund universe.[5] We find that about 14.8% of filings in our sample show signs of significant portfolio adjustment prior to reporting and display a pattern of returns consistent with window dressing activity. About 5.5% of funds repeatedly report portfolios that are misleading. Furthermore, the turnover generated by this trading activity does not, on average, result in abnormal returns after the reporting date. The funds identified as window dressers are typically growth funds with high turnover, high expense ratios, large cash holdings, and a poor recent performance. Moreover, portfolios filed by funds that subsequently died were more likely to have been window dressed. Regardless of the overall performance of a fund, substantial trading over a short period of time before reporting to window dress the portfolio holdings will incur immediacy costs without providing any added value. This is against the interests of shareholders and raises questions about the governance of mutual funds. There is an ongoing debate on whether mutual funds should legally be required to report more frequently. Given that holdings do not necessarily provide more insight into the risk and performance of funds, we conjecture that it would be much more effective if funds were required to

[4] Ritter (1988) and Dyl and Maberly (1992) also investigate the "turn-of-the-year" effect, and Brown, Harlow and Starks (1996) analyze fund activity to manipulate risk.

[5] "Mutual Fund Fact Book", *Investment Company Institute* (ICI), May 2003, p.68. The net assets of world equity funds ($358.5 billion) are subtracted from the total for equity funds.

disclose the trading date of, say, its top ten acquisitions and sells. Such disclosures could be delayed to avoid imitation of successful funds. This would enable the investor to get a better understanding of the selection ability of the fund manager as well as the risk characteristics and performance of past returns.

We compare our findings with a number of competing explanations including liquidity costs, end-of-year effect, and momentum trading. None of these competing hypotheses is able to explain the observed return differences around the reporting date. When comparing with NAV returns we add back management fees and asset-based costs for distributing the fund's shares. To control for invisible execution costs, we estimate the price impact of trading in the stocks that the fund holds. We find that liquidity costs cannot explain the magnitude of the differences in returns before the reporting date. Moreover, if liquidity costs were the prime reason for the diverging return differences, we would expect to see the same pattern after the portfolio date. This is not what we find in the data.

We find some support that filings in December are more likely to be window dressed. However, window dressed portfolios are not clustered in December and the pattern for the filings in June or December do not exhibit different patterns and magnitudes of underperformance. We conclude that tax-loss selling cannot explain the asymmetries we find in the return differences.

A classical momentum trader, along the definitions in Jegadeesh and Titman (1993, 2001), routinely adds recent winners to his portfolio and sells losers. This type of strategy will, by construction, imply that the NAV returns always underperform the reported end of quarter portfolios. However, we show that momentum trading in general leads to a gradual convergence to zero of the return discrepancy as we approach the reporting date. This is in contrast to the pattern observed in the data for window dressed filings, where the return difference remains large until the last days of the quarter.

Our finding is also different from the leaning-for-the-tape argument of Carhart, Kaniel, Musto and Reed (2003). They find evidence that funds bid up prices of stocks they already hold in their portfolio immediately before the closing of the stock exchange on the last trading day before disclosing their holdings. This is a way to manipulate the performance and can be done by placing relatively small orders. The authors observe abnormal returns over the last half hour of the quarter and a reversal on the next trading day. Unless the orders are large and induce major shifts in the portfolio compositions, this behavior would equally affect prices for the reported portfolio as well as the NAV returns. Hence, our identification method is not distorted by this activity.

The remainder of the paper is structured as follows. Section 2 describes the sample. Section 3 introduces a new methodology to identify reported holdings that are window dressed. Section 4 discusses the typical characteristics of these portfolios and examines the persistence of window dressing activity at the fund level. Section 5 compares our identification method with an analysis of net changes between two adjacent portfolio dates and Section 6 concludes.

## 2 Data Description

Our data set of U.S. domestic equity, open-end funds covers semi-annual holdings from 1997 to 2002.[6] Almost all funds in our sample report at quarterly frequency, but our database covers only semi-annual filings.[7] For each fund we observe the daily net asset values (NAV) per fund share along with a track record of all distributions. The data is provided by Morningstar, Inc. (hereafter Morningstar). The sample of holdings data contains 27,702 reported portfolios from 4,025 funds. We match each reported stock position with daily return data from the Center for Research in Security Prices (CRSP). Filings where we can match at least 95% of all positions (in terms of market capitalization) form the subset used to identify window dressers. This cutoff reduces the sample to 3,289 funds with 18,139 filings containing a total of 10,607 different securities. The Investment Company Institute (ICI) counts 4,756 equity funds at the end of 2002, including world equity funds. The total market capitalization of the funds in our sample is $1.950 trillion at the end of 2001, which accounts for 54.5% of the U.S. domestic equity fund industry. A total of 209 funds with 1,074 filings in the subset are index funds. The sample also includes 16.9% dead funds and 10.8% of the filings are reported by these funds.

To control for invisible trading costs we calculate a measure of price impact using intraday data from the Trade and Quote (TAQ) database. Closely following the methodology proposed in Breen, Hodrick and Korajczyk (2002) we relate returns and net trades over half-hour periods to estimate the price impact of trading individual stocks. At the fund level, the market capitalization weighted average measures the price impact for trading 1 million of the underlying fund portfolio within half an hour. The details of these calculations are described in Appendix A. The 95% matching criterion assures that we can assign a liquidity measure for the fund's holdings with a reasonable degree of accuracy.

### 2.1 The Subsample Used to Identify Window Dressers

Panel (a) in Figure 1 summarizes the fraction of all portfolio positions that can be matched with daily returns from CRSP for all 27,702 filings from 4,025 funds. To be included in the subsample used to identify window dressers, at least 95% (market capitalization weighted) of the positions must be matched. The cutoff at 95% is indicated by the solid, vertical line. Panel (b) describes the number of filings per fund for the subset of 3,289 funds (and 18,139 reported portfolios) where we can assign daily returns for at least 95% of the net assets. At most, there are eleven semi-annual filings between June 1997 and June 2002. For 1,822 funds 5-11 filings are observed, which in total account for 80.6% of the

[6] With 36.1% of the overall fund industry assets, this is the largest fund category, closely followed by money market funds with 35.5%.

[7] The reporting date to Morningstar may differ in some instances from the reporting date in the SEC's EDGAR database but we do not find that non-EDGAR filings differ from EDGAR filings in our analysis.

reportings used to identify window dressers. For 2,168 filings the full set of the reported securities can be linked with a daily returns. On average the sample contains 5.5 filings per fund.

Table 1 compares the number of funds and the market capitalization of the subset of 3,289 funds with 95% matching positions to the full sample of 4,025 U.S. domestic equity funds. The table illustrates that the sample of these selected funds accounts for a representative number of all U.S. equity funds within each of the nine Morningstar style orientations.

## 2.2 Fund Characteristics and Investment Style

Fund characteristics such as the expense ratio, turnover, investment style, rating, etc., are provided by Morningstar at annual frequency. Missing values for fund characteristics, in particular expense and turnover ratios, were complemented with data from the CRSP Survivorship-Bias Free US Mutual Fund Database (MFDB). Using intraday trades and quotes on the underlying individual stocks from the Trade and Quote database (TAQ), we assign a liquidity measure to each reported portfolio as the weighted average of the liquidities of the individual stocks in the portfolio.

### 2.2.1 Fund Styles

Funds are often classified into investment styles according to value-growth and large-small cap characteristics. We use Morningstar's assignments to the nine quadrants of the style box corresponding to the three size classes large, medium and small cap and the three book-to-market classifications value, blend and growth.[8] Morningstar defines large stocks as the group of stocks with the largest market capitalization that contributes 70% to the total market capitalization of all publicly traded, domestic stocks. The next 20% are medium size stocks and the remaining 10% small-cap stocks. The geometric mean of all stocks held by the fund determines its position on the vertical axis.[9]

The horizontal axis of the style box is split into the categories value, blend, and growth. On the individual stock level, Morningstar calculates a value and growth score based on historical and prospective stock characteristics. The historical and forward-looking data each contribute 50% to the scores. The historical data components for the value score are price-to-book, price-to-sales, price-to-cash flow and dividend yield; each with equal weight. Growth score components are historical earnings growth, sales growth, cash flow growth, and book-to-value growth. The market capitalization weighted value/growth score of the individual positions determines the fund style.[10]

[8] An alternative would be to classify funds using Sharpe's (1988, 1992) style regressions.

[9] Suppose the fund holds three stocks A, B and C with weights 60%, 30% and 10% with market capitalizations of 100, 10 and 1,000 million. The geometric mean is then $(100^{0.6}) \times (10^{0.3}) \times (1000^{0.1}) = 125.90$, which is larger than the median market capitalization, but below the mean of 263.00.

[10] For a detailed description of the Morningstar style box classification see www.morningstar.com, "Mutual Funds", help

### 2.2.2 Fund Characteristics Vary with Style

Table 2 breaks the average fund characteristics down by style. The list covers the characteristics that we consider to be potentially significantly different between window dressers and non-window dressers. Morningstar rates funds from five stars to one star based on past performance. Risk-adjusted fund returns are evaluated relative to peers within the same investment style.[11] A pre-specified fraction of all funds receive five to one stars: The top 10% get five stars, the next 22.5% four, the middle 35% three, 22.5% two, and the bottom 10% one. Funds in existence for less than three years are not rated (rating 0).

Index funds with a large number of stocks explain the discrepancy between the arithmetic average number of stocks and the median values. The broadest fund in our sample, the Fidelity Spartan Extended Market Index (ticker FSEMX) held 3,468 stock in June of 2001 and is a member of the medium blend category. Giants like the Vanguard 500 Index fund ($65.9bn in December 2002) and the biggest actively managed fund, Fidelity Magellan ($62.5bn), dominate the average market capitalization of large blend funds. Cash holdings do not differ systematically across investment styles. The liquidity costs are measured by the estimated price impact (in basis points) when shares worth $1 million of the underlying portfolio are traded over a half hour period. When compared to the highly liquid large-cap universe, the price impact is approximately five times larger for medium-cap, and more than 20 times larger for small-cap funds. Average turnover ratios and expense ratios are larger for growth funds. As mentioned above, index funds constitute an important fraction in the large and medium blend categories.

The distributions of these characteristics are shown in Figure 2. In Panel (a), S&P 500 index funds are the main reason for the second mode to the right in the distribution of number of stocks in the portfolio [$\log_{10}(500) = 2.7$]. A total of 464 filings list between 490 and 510 different equity positions.[12] It is interesting to note that for equity funds 5% cash seems to be a critical threshold [Panel (c)]. Only 5.7% of all funds report positions in cash or cash equivalent securities beyond 5% of total net assets. The average fraction of total net assets that is invested in the top 10 holdings is 34.4% [Panel (d)]. The left tales in the distributions of expense ratio and annual turnover [Panels (e) and (f)] are again mainly due to index funds. A large fraction (43.6%) of the filings is from funds which are not rated by Morningstar [Panel (g)]. On average, fund managers have been with the same fund for 4.6 years as of

function for “Style Box”.

[11] Morningstar enhanced the rating system in July 2002. The new method uses a refined grid of peer groups and an adapted risk-adjustment to put more emphasis on downward variation [“Fact Sheet: The New Morningstar Rating for Funds”, www.morningstar.com, 2002]

[12] These filings are reported by 98 different funds, of which 76 have the word “Index” in its name. In 32 cases the name contains explicitly “S&P 500 Index”.

the filing date [Panel (h)].

# 3 Identification of Return Window Dressing

We propose a methodology to identify return window dressing, defined as the case where a fund manager reports securities with a relatively high recent return in order to convey superior stock picking ability to the investor. Prior to disclosure, a window dresser would replace stocks in his portfolio that had a poor recent return with stocks that performed well. The return characteristics of the reported securities are then no longer representative of the portfolio actually held by the fund throughout the quarter.

The traditional approach to analyze trading activity by fund managers has been to look at the net changes in portfolio positions between subsequent filings. The main shortcoming of this approach is that the comparison of subsequent reported portfolios does not reveal any information about the timing of these transactions. Furthermore, little is revealed about the total trading activity during the quarter for funds with high turnover since only net turnover can be measured. Determining the timing of trades is crucial for detecting window dressers whose distinguishing trait is that they will load up on quarter winners and/or dispense with poorly performing stocks on their books immediately before the reporting date

While window dressing thus cannot be detected using holdings information alone, more can be learned by also exploiting daily data on fund net asset values and individual stock prices. In particular, such trading activity will result in a divergence of the return on the reported portfolio and the observed fund return. The shape test proposed in this paper is designed to identify the distinct return pattern associated with a window dressed portfolio.

## 3.1 Buy-and-Hold Returns

Any test of window dressing attempts to assess how representative the disclosed portfolio is of the actual portfolio held during the quarter, controlling for the average level of turnover of the fund. To achieve this goal, we determine the hypothetical return the fund would have earned, if it had held the reported portfolio over the weeks prior to reporting at quarter end. These buy-and-hold returns are the benchmarks against which we compare the fund's realized returns in order to identify window dressing. Determining the total holding period return for the buy-and-hold strategy requires taking into account distributions and share splits, the calculation of which is described in this subsection.

The reporting dates in our sample are the two available filings closest to June 30th and December 31st each year and for the majority of the funds the reporting date is at the end of these two quarters. The portfolio holdings are at the individual fund level and show the complete record of holdings of stocks, bonds and other assets. For each stock position in the mutual fund portfolio we retrieve daily stock prices from CRSP to create a time series of the portfolio returns starting 91 days before the reporting date and ending 91 days afterwards. The portfolio weight of each stock is calculated as the fraction of total fund assets held in that stock. We define the fund's cash position as the dollar value of all cash and cash equivalent instruments. Filings where less than 95% of the total assets less cash can be matched with daily returns from CRSP are excluded from the sample.[13]

Let time $t = 0$ denote the reporting date. The buy-and-hold portfolio weights for stock $i$ are then, for any $t = -91, ..., 91$, defined as

$$w_{i,t} = \frac{P_{i,t} n^0_{i,t}}{\sum P_{i,t} n^0_{i,t}} \tag{1}$$

where $P_{i,t}$ is the closing price or the average of the bid and ask price at the end of the trading day, and $n^0_{i,t}$ the adjusted number of shares for stock $i$ relative to date $t = 0$. To account for stock splits, reverse splits, right issues, share buybacks, spin-offs, issuances and offers, we adjust the number of shares at different points in time. The reference point for adjusting the number of shares is always the reporting date. Take a 2-for-1 split as an example and consider the two cases where (i) the stock split occurs before the reporting date, and (ii) after the reporting date. The adjustment factor is $Q_i = 1$ before and $Q_i = 2$ after the event. In general,

$$n^0_{i,t} = n_{i.t} \frac{Q_{i,0}}{Q_{i,t}}$$

where $n_{i,t}$ denotes the unadjusted number of stocks. In scenario (i) $n_{i,t}$ is adjusted by $Q_{i,t}/Q_{i,0} = 1/2$, whereas in scenario (ii) the appropriate adjustment is $Q_{i,t}/Q_{i,0} = 2/1$.[14]

The return on the buy-and-hold strategy for a reported fund portfolio, $R^P_t$, is computed as

$$R^P_t = \sum_i w_{i,t} r_{i,t} \tag{2}$$

where $w_{i,t}$ are the portfolio weights as defined in (1) and $r_{i,t}$ are the corresponding holding period returns from CRSP.[15]

[13] Most funds and reportings where less than 95% is matched had large Treasury bond, corporate bond, or international stock positions.

[14] We use the cumulative factor to adjust number of shares, CFACSHR, from CRSP. Instead of using the differences in adjusted prices we use the holding period returns between two dates.

[15] We assign a return of zero to cash and cash equivalent positions. This underestimates the return on the buy-and-hold strategy and reduces the underperformance of NAV based returns. Note that such a bias is very small over the holding periods considered below (e.g. weekly returns) and would in any case tend to reduce the number of window dressed portfolios we can detect. Stock positions that we cannot match with a daily return series are also assumed to have zero return. On average, this will likely underestimate the return of the buy-and-hold strategy as well. An alternative would be to assign the average return of the rest of the holdings as the return of the unmatched position. To ascertain whether our 95% requirement for the fraction of matching positions does not introduce any systematic bias we recalculated all the main results with the subset of fund portfolios with 100% matching positions and did not find any qualitative differences.

### 3.2 Difference Between Holdings-Based and NAV Returns

By investing in a fund the investor benefits from professional management of the portfolio. Independent of the success of the trading activity the fund provides services such as collecting and reinvesting dividends, keeping track records of shareholders, producing reports on performance, etc. The financial adviser who manages the fund and any affiliated service providers are compensated through management fees and service fees. To allow a fair comparison with a buy-and-hold strategy we add back these operating expenses to determine the fund's pre-expense performance. Fees that are not paid out of the fund's assets but are rather directly paid by the investor do not affect this adjustment. The details of the adjustment calculations are given in Appendix A.

Figure 3 and 4 show the pattern of average daily return differences for the sample of 18,139 filings from 3,289 funds for which at least 95% of the holdings can be matched with daily returns from CRSP. The return difference is defined as the average daily return on the buy-and-hold strategy minus the realized NAV return, $R_t^P - R_t^{NAV}$. When the hypothetical returns on the portfolio that is reported at the end of the quarter outperform the fund's realized returns, the difference is positive. The distributions of return differences shown in Figure 4 exhibit the expected leptokurtic shape typical of daily stock returns and are centered at zero with a dispersion that increases with the distance in time to the reporting date, consistent with the average manager having no skill. After the reporting date, we do not detect any distinct asymmetry around zero, but over the last four weeks prior to the reporting date there is a clear right skew indicating that a significant number of managers report portfolios which outperformed the funds' NAV based returns. This is of interest because we would precisely expect a return window dresser to show up in the right tail of the distribution prior to reporting. This, of course, need not be the case for other types of window dressing, e.g. for a style rotator who took a successful bet on another style during the quarter.

### 3.3 A Shape Test to Identify Window Dressing

For a manager who buys recently successful stocks and removes losers from his portfolio just before disclosure, the buy-and-hold strategy of the ex-post reported portfolio will outperform the NAV returns over the weeks prior to the reporting date, thus leading to a significant positive return difference. A key feature that distinguishes a window dresser from, a momentum trader is the timing of trades. While the window dresser concentrates a large fraction of his turnover just prior to reporting, the momentum trader adds recently successful stocks on an ongoing basis. To illustrate the differential impact of these alternative trading patterns, we conduct a small scale Monte Carlo experiment, the results of which are displayed in Figure 6. In the experiment, we consider three hypothetical momentum traders who are assumed to turn over 5% of their portfolios each week based on stock performance over the previous 1, 2 or 4 quarters respectively. As can be seen from the figure, this rebalancing strategy leads to a large positive return difference throughout the quarters leading up to the reporting date, followed by a gradual convergence of the return differences to zero. Since the simulated stock returns are assumed to be serially uncorrelated, the momentum strategies are necessarily unsuccessful ex-post, as can be seen from the zero average return difference following the reporting date. For a successful momentum trader, the only difference would be a negative return difference post reporting, indicating that the momentum strategy outperforms the buy-and-hold strategy.

By contrast, the window dressers concentrate their trades in the week immediately before reporting. We assume that the first window dresser turns over 20% of his portfolio by selling the quarter losers on his book and buying a subset of the quarter winners. This leads to a very characteristic pattern: one quarter prior to reporting, the return difference increases sharply from its previous level of zero and remains high before dropping off sharply at the reporting date. After the reporting date, there is no systematic difference between the NAV based returns and the returns on the buy-and-hold strategy in the absence of manager skill. The period over which the return difference is positive equals the horizon based on which the window dresser determines winners/losers in this simple example. To see this, compare the profile of the window dresser who has a 4 week horizon, also shown in Figure 6. For this window dresser, the return difference is positive over the 4 weeks leading up to the reporting date and zero otherwise. We also see that the shorter the window dressers horizon for picking stocks, the larger the return difference. Similarly, the return difference will increase with turnover due to window dressing.

More generally, window dressers may of course choose different horizons or a combination of horizons. This will result in a pattern of return differences equal to a weighted average of return patterns of the type shown for the window dressers in Figure 6. Importantly, a common feature of all window dressers is that the return difference is substantial (and much larger than for a typical momentum trader) over the last few weeks before reporting.

Our identification of window dressed portfolios is inspired by this observation. It proceeds by comparing the realized daily net asset value returns of fund $i$, $R_i^{NAV}$, with the daily returns on the buy-and-hold strategy of the reported portfolio, $R_i^P$. We take the reporting date to be time zero and consider the 4 weeks leading up to and following this date, denoted by $w = -4, ..., -1, 1, ..., 4$. For each reported fund portfolio, $i = 1, ..., N$, and each week, $w$, we calculate the difference between the average daily return on the buy-and-hold strategy and the realized NAV returns, $\Delta_{i,w} = \bar{R}_{i,w}^P - \bar{R}_{i,w}^{NAV}$. The goal is to distinguish window dressers from both momentum traders as well as the average manager who has a zero return difference throughout (in absence of skill) based on a significant positive average $\Delta_{i,w}$

over the last four weeks of the quarter. This means that window dressed portfolios can be identified provided that the window dresser's turnover results in a significant return difference, i.e. that the window dresser mainly chooses stocks based on their performance over the past 4-13 weeks and that a significant fraction of his total turnover occurs over the last couple of weeks of the quarter. If the average horizon of the window dresser is much longer than one quarter or the turnover over the last weeks is small, we will not be able to identify the portfolio as window dressed.

We make the following assumptions which will allow us to write up a simple parametric model for the weekly return differences $\{\Delta_{i,w}\}_{w=-4,\dots,4}$. The assumption essentially says that, controlling for turnover, the only thing that distinguishes window dressers, momentum traders, and the average manager is the pattern of average return differences while the 'dispersion' is the same.

**Assumptions***:*

(i) (GED) *The distribution of* $\Delta_{i,w}$ *can be approximated by a Generalized Error Distribution (GED) with shape and scale parameters which depend only on* $w$ *and the annual turnover of the fund, but not on* $i$.

(ii) (Independence) *The weekly return differences,* $\Delta_{i,-4},\dots,\Delta_{i,4}$*, are independent over time.*

(iii) (Symmetry) *The shape and scale parameters of* $\Delta_{\cdot,w}$ *are identical to those of* $\Delta_{\cdot,-w}$.

(iv) (No average skill) *In absence of window dressing,* $E[\Delta_{i,-w}] = E[\Delta_{i,w}] = 0$.

A few comments are in order. Due to the weekly averaging, there are only four observations associated with each filing date which are used for identifying window dressing. This results in the need for a tight parametrization in order to retain identifying power. The choice of the generalized error distribution is somewhat arbitrary, but it is often used in the GARCH literature because it is a simple generalization of the Gaussian distribution which is able to capture the leptokurtic features prevalent in daily returns data.[16] All we need for our purposes is a parametrization which fits the observed marginal distributions reasonably well without too many free parameters to be estimated (in the case of the GED there are two parameters). Figure 4, Panels (a)-(d) illustrate how well the GED fits the marginal distribution of return differences over the weeks after the reporting date where we expect the average return difference to be zero. The independence assumption is crucial, and would likely be a bad assumption if we were analyzing daily return differences directly. Instead, we choose to construct weekly averages in order to limit possible GARCH effects in the data. While this simplifies the analysis considerably, the downside is that it leaves us with fewer observations per filing (i.e. one per week). Clearly, the distribution of $\Delta_{i,w}$ will depend on $w$, the time from the reporting date where, by definition, $\Delta_{i,0} \equiv 0$

[16] See for example Nelson (1991).

must hold. In particular, we would expect the variance of $\Delta_{i,w}$ to increase for larger $w$ due to the continuing turnover of the funds, as can be seen from Figure 3. In absence of skill or window dressing, however, there is no reason why the distribution of $\Delta_{i,w}$ should depend on the sign of $w$. Implicit in this assumption is that any (right) skewness in the distribution of return differences prior to reporting is due to window dressing by fund managers while a left skew post reporting would be due to managers with skill. No such left skew is detected in Figure 4, which is consistent with the average manager not being able to beat the buy-and-hold strategy before fees. That there is a tendency to report portfolios which have performed relatively well can be seen from Figure 4, Panels (e)-(h), which display the marginal distribution of the return difference $\Delta_{i,w}$ over the weeks prior to reporting. Comparing with the distribution post reporting shown in Panels (a)-(d), the empirical distributions pre-reporting are clearly right skewed over the last 4 weeks. Going further back than 4 weeks, the asymmetry largely disappears. This observation motivates our choice of the 4 week period prior to reporting for identifying potentially window dressed portfolios.[17]

**Hypothesis**: Window dressed portfolio $i$:

*The distribution of* $\Delta_{i,w}$, *satisfies*

$$E[\Delta_{i,w}] = \quad \theta_i > 0 \quad \text{if} \quad -4 \leq w \leq -1$$

The identification of window dressed portfolios now proceeds by identifying fund filings $i$ for which $E[\Delta_{i,w}]$ is positive and significant over the 4 weeks leading up to the reporting date. In order to control for turnover, we first divide the funds into groups based on their annual turnover such that funds with an annual turnover of $0-25\%$ constitute the first group, $25-50\%$ the second group, etc. In the first step of the analysis, we establish the distribution of $\{\Delta_{i,-w}\}_{w=1,\ldots,4}$ conditional on turnover $\tau$ under the null of no window dressing. The symmetry (in $w$) and no average skill assumptions imply that the distribution of $\Delta_{i,-w}$ is the same as that of $\Delta_{i,w}$ under the null of no window dressing. Moreover, the latter is not affected by the presence of window dressed portfolios and therefore a consistent estimate of the distribution of the pre-reporting return differences under the null can be obtained from the post-reporting return differences.

For each turnover group, $\tau$, and each week $w = 1, \ldots, 4$ after the reporting date, we fit a GED distribution to the return differences $\Delta_{i,w}$ of fund filings $i$ belonging to turnover group $\tau$ using maximum likelihood estimation:

$$(\eta_w^\tau, \sigma_w^\tau) = \arg\max_{\sigma,\eta} \sum_{\{i | \text{ fund portfolio } i \text{ has turnover } \tau\}} \log f\left(\Delta_{i,w} \mid \eta, \sigma\right) \tag{3}$$

[17] While momentum trading also can lead to a right skew prior to reporting, this effect should be relatively small based on our simulation experiment reported in Figure 6. In Section 4 below we explicitly test whether momentum traders mistakenly are picked up by the shape test.

where the GED density $f(.)$ is given by

$$\log f(x,\eta,\sigma) = \left\{\log\left(\frac{\eta}{\lambda}\right) - \left(1+\eta^{-1}\right)\log(2) - \log\left[\Gamma\left(\eta^{-1}\right)\right]\right\} -0.5\left|\frac{x}{\lambda\sigma}\right|^{\eta} - \log(\sigma)$$

and $\Gamma(.)$ is the gamma function and we have defined

$$\lambda = \left[\frac{2^{-2/\eta}\Gamma(1/\eta)}{\Gamma(3/\eta)}\right]^{0.5}.$$

The resulting set of parameter estimates $(\hat{\eta}_w^\tau, \hat{\sigma}_w^\tau)$ for each turnover category and each week $w$, are displayed in Figure 5. The parameter $\sigma_w^\tau$ is a scale parameter while $\eta_w^\tau$ governs the degree of non-normality of the distribution. When $\eta_w = 2$, the distribution is a normal, while $\eta_w < 2$ implies leptokurtic features. As can be seen from Figure 5, non-normality is clearly an important feature of the data.

According to the symmetry assumption, the estimated distributions of post-reporting return differences are consistent estimates of the distributions of the pre-reporting return differences under the null, i.e. we can take $\eta_{-w}^\tau = \eta_w^\tau$ and $\sigma_{-w}^\tau = \sigma_w^\tau$ for $w = 1,\ldots,4$. In what follows we take these first-step estimates as *given*, such that for each fund filing $i$, we can formulate the test of no window dressing against the (one-sided) alternative of window dressing as a simple hypothesis about the mean of the return differences:

$$H_0 : E[\Delta_{i,-w}] = 0 \qquad \text{against} \qquad H_A : E[\Delta_{i,-w}] = \theta_i > 0 \quad \text{for} \quad w = 1,\ldots,4 \tag{4}$$

where $\Delta_{i,-w}$ is i.i.d. GED with unknown mean $\theta_i \geq 0$ and given scale and shape parameters $(\eta_{-w}^\tau, \sigma_{-w}^\tau) \equiv (\hat{\eta}_w^\tau, \hat{\sigma}_w^\tau)$. Filing $i$ is then identified as potentially window dressed if $H_0$ can be rejected at the 5% level of significance (respectively 1% level).

Two subtle issues arise in testing (4). First, and perhaps least obvious, the null hypothesis is not in the interior of the parameter set. This causes problems for the classical maximum likelihood theory because the first order conditions need not hold at the maximum likelihood estimator and the standard limiting distributions will not be valid in general.[18] Instead, we follow Andrews (2001) in computing

[18] King and Wu (1997) advocate the use of a locally most powerful test based on the statistic

$$U_i = \sum_w \frac{\partial}{\partial\theta}\log f(\Delta_{i,w} - \theta, \eta_w, \sigma_w)\bigg|_{\theta=0}$$

Intuitively, $U_i$ will be small when there is little scope for increasing the likelihood function by a marginal increase in $\theta$, and thus the test of (4) should reject for fund filings with large (positive) values of $U_i$. The asymptotic distribution of $\sqrt{n}U_i$ can be shown to be normal under the null. Unfortunately, we found the asymptotic test to have very limited power against non-local alternatives in our small sample setting.

the Quasi Likelihood Ratio (QLR) defined as

$$QLR_i = -2\left\{\left[\sum_{w=-4}^{-1}\log f\left(\Delta_{i,w},\eta_w^{\tau},\sigma_w^{\tau}\right)\right] - \sup_{\theta\geq 0}\left[\sum_{w=-4}^{-1}\log f\left(\Delta_{i,w}-\theta,\eta_w^{\tau},\sigma_w^{\tau}\right)\right]\right\} \qquad (5)$$

the limiting distribution of which will not in general be $\chi^2$. The second issue encountered is that only 4 observations are available to identify $\theta_i$ for each fund filing and the distributions involved are highly non-normal. This means that the asymptotic distribution theory in Andrews (2001) will likely provide a poor approximation of the finite sample distribution of $QLR_i$. For this reason we instead rely on Monte Carlo simulations to provide exact finite sample critical values for $QLR_i$, taking the first step estimates of $(\eta_{-w}^{\tau},\sigma_{-w}^{\tau})$ as given. Furthermore, we compute the power function of the test displayed in Figure 7. It shows the rejection frequency of $H_0$ as a function of the true $\theta_i$ for each turnover category (i.e. the probability of incorrectly categorizing a window dressed portfolio as non-window dressed). The test retains the ability to identify return differences in excess of 5-10 basis points per day for all but the highest turnover categories. The power does suffer when turnover is high which, all else equal, will limit our ability to identify window dressed portfolios in the highest turnover categories.

Based on the test (5), we identify 2,678 fund filings, corresponding to 14.8% of our sample, for which the pattern of return differences characteristic of a window dresser is identified at the 5% level of significance. At the 1% level of significance we find 1,021 fund filings to be window dressed, corresponding to 7.1% of the sample.

### 3.4 Average Return Differences for Window Dressers

To illustrate the scale of the return differences generated by window dressing we repeat the boxplot from Section 3.2 for the 2,678 window dressed (WD) portfolios and the non-WD portfolios separately. Figure 8 shows the return differences between the returns on a buy-and-hold strategy of the reported portfolio and the realized NAV returns for non-overlapping, weekly intervals before and after the reporting date, $\Delta_w = \overline{R}_w^P - \overline{R}_w^{NAV}$. The gray shaded boxes measure the 25th and 75th percentile, and the line drawn across the median. Compared to Figure 3 the T-bars indicating the range where 95% of the observations fall into are dropped. Panel (b) describes the subset of 2,678 portfolios that are identified as window dressed according to the shape test.

The median return difference is approximately 5 basis points per day during the quarter leading up to the reporting date. The weekly intervals are non-overlapping periods and it is not a priori given that geometric mean return differences would on average remain positive each week. The drop in these return differences over the last week is substantial and the interquartile range for all 2,678 portfolios is clearly shifted upwards. Most strikingly, there is no evident pattern after the reporting date. There no longer seem to be any systematic deviations from zero. Simply holding the reported portfolio over the next quarter (13 weeks) would, for the average fund, yield the same return. What makes this comparison even more striking is that it is based on NAV returns before operating expenses are deducted. Panel (a) contains the return differences for the subset of the remaining set of funds that are not classified as window dressers. The return differences for these funds still show a symmetric pattern around the reporting date as in Figure 3 while the medians are close to zero and the dispersion increases the further away from the reporting date.

In the case of index fund we of course do not expect to see any systematic difference between the returns on the reported portfolio and the realized fund returns around the reporting date. Furthermore, to the extent that index funds have lower turnover than non-index funds, we expect to see a smaller dispersion of the return differences. Figure 8 Panels (c) and (d) show the pattern of return differences for non-index and index funds respectively.

# 4 Characteristics of Window Dressers

What characteristics do we expect to find for window dressed portfolios? The preceding section introduced an identification method to detect window dressing activity based on daily NAVs as well as reported portfolio holdings. This section investigates whether the identified portfolios indeed exhibit the characteristics one would expect from funds engaging in window dressing. This provides an important validation to rule out the possibility that they were identified by chance.

To be specific, we would expect window dressers to possess some or all of the following properties:

(i) Medium to high turnover. The example later in this section illustrates that the spread between the realized fund return and the buy-and-hold return that we observe for window dressed portfolios easily translates into high turnover.

(ii) Funds with poor past performance are more tempted to window dress. A downside revision of the fund rating can induce investors to withdraw capital allocated to the fund. Del Guercio and Tkac (2003) find significant abnormal in- and outflows due to rating changes. Current rating systems most often incorporate holdings information.

(iii) For funds with a short history, the investor may pay more attention to holdings given the limited return series.

(iv) Investors who wish to see "glamour" stocks in a fund portfolio will be more likely investing in growth stocks. Value investors are considered to be more hard-nosed.

(v) Funds holding liquid stocks may engage more often in window dressing activity as trading in these stocks is less costly.

(vi) Some degree of persistence. A fund manager who did window dress is more likely to do it again. He will not necessarily adjust his portfolio before every reporting date but whenever his stock picking turns out to be unsatisfactory.

The results of the probit analysis below shed light on the link with these characteristics. A stylized example illustrates that the additional turnover generated by window dressing activity is substantial. We also address the question whether the window dressed filings are clustered over time or in December. The medium return difference we observe for window dressers over the last month prior to reporting is approximately five basis points per day. To investigate whether these differences can be explained by invisible trading costs we estimate the price impact of turning over the fund portfolio. We find that these liquidity costs cannot explain the order of magnitudes we observe.

## 4.1 Probit Results

To examine the relationship between fund characteristics and window dressing, we conduct a probit analysis in which we assume that the conditional probability of a portfolio being window dressed (WD) is related to a linear function of certain characteristics:

$$P(y_i = 1|x_i) = \Phi(x_i\beta)$$

where $\Phi(\cdot)$ is the Gaussian CDF. The 0-1 variable $y_i$ indicates whether the shape test found filing $i$ to be window dressed ($y_i = 1$) and $x_i$ is the set of characteristics we wish to relate to window dressing activity. The characteristics include fund size, number of stocks in the portfolio, cash holdings, the turnover over the calendar year, and past performance. The expense ratio measures the explicit management costs and the liquidity of the holdings is quantified by the average price impact. Five dummy variables indicate whether the fund follows a momentum strategy, whether the filing is at the end of December, in an up-market, whether it is an index fund, and whether it is a dead fund. The remaining explanatory variables control for the Morningstar rating and the nine style orientations. We find that turnover, past poor performance relative to peers, expense ratio, style orientation, cash holdings, the number of stocks, the size of the fund and whether the fund is dead are the most persistent factors.

The results are summarized in Table 3. Included at the top are the number of observations and the pseudo R-squared. The number of observations is reduced to 11,834 as we exclude filings that are not assigned to a stylebox. The models I-5 and I-1 contain the full set of explanatory variables with $y_i$ determined by the shape test at a level of significance of 5% and 1% respectively. The models II-5 and II-1 are defined similarly except that insignificant characteristics (at the 5% level) have been removed by successive testing. As can be seen from the row showing the pseudo R-squared, these values do not drop much by the removal of insignificant characteristics.

We find that funds with considerable cash positions and fewer stocks in the portfolio are more likely to window dress. This appears to be plausible as holding cash and a low number of stocks facilitate window dressing. Fund size, measured as the market capitalization of the portfolio, is no longer an important determinant once we use the WD portfolios at the 1% significance level. Next, we consider turnover. The construction of the shape test controls for the annual turnover and, thus, does not imply that high turnover funds automatically are more likely to be classified as window dressers. Still, the coefficient for turnover is significant and confirms that highly actively trading funds more often tend to window dress. A stylized example in the next section shows how window dressing can easily create a substantial increase in annual turnover.

Poor past performance over the quarter leading up to the reporting date is measured as the cumulative excess return over the quarter relative to the peers within one of the nine styleboxes. The results confirm that funds performing poorly throughout the quarter leading up to the reporting date window dress more often. Conversely, as the coefficient in the table shows, the higher the excess performance the lower the probability that the portfolio will be window dressed. This is also consistent with the finding that funds that eventually ended up as dead funds in our sample are more likely to have window dressed.

Our shape test is designed to distinguish window dressers from momentum traders. Another way to cross-check whether the funds that the shape test identifies as window dressers are not funds that follow a momentum strategy is to regress the returns on the Fama/French benchmarks and a momentum factor. For a given fund, for which at least one filing is classified as WD, we run the regression over all past monthly returns that are available. The regression is

$$r_{i,t} = \alpha_i + \beta_{1,i}\left(r_{m,t} - r_{f,t}\right) + \beta_{2,i} SMB_t + \beta_{3,i} HML_t + \beta_{4,i} UMD_t + \varepsilon_{i,t} \tag{6}$$

where $r_{m,t} - r_{f,t}$ is the excess return of the market over the risk-free rate, $SMB$ denotes the performance of small stocks relative to big stocks, $HML$ the performance of value stocks relative to growth stocks, and $UMD$ the momentum factor.[19] The momentum factor is calculated as the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios. Therefore, if the funds that we identify as window dressers are indeed following a momentum strategy we would see on average a positive coefficient for the momentum factor. A dummy variable indicating whether the coefficient on momentum is significant or not is included as an additional explanatory variable in the probit model. There appears to be no evidence of WD filings loading up more on the momentum factor than non-WD filings. We therefore conclude that the filings identified as window dressed are not due to a mis-classification of momentum strategies.

[19] The data is taken from Kenneth French's website.

Another dummy variable equals one if the filing is reported in December. The coefficient is significant and indicates that window dressing occurring, all else equal, is more likely in December. Similarly, it is more likely that a fund window dresses in a bull market. This raises the question as to whether the portfolios that are identified by the shape test as likely window dressed are clustered over time. Table 4 tabulates the occurrence of the WD portfolios over time and in December versus non-December months. The table provides evidence that this is not the case. Repeating the boxplot from the last section for December and non-December WD fillings, plotting the average daily return differences between the buy-and-hold return of the reported portfolio and the realized NAV return over weekly intervals, does not exhibit any qualitative differences. Panels (a) and (b) of Figure 9 break the 2,678 WD portfolios down into two subsets; one with all 1,610 WD portfolios in months other than December, and another with the 1,068 WD portfolios with reporting dates at the end of December. The comparable number of filings as well as the patterns of the return differences rule out the hypothesis that the window dressing activity is primarily driven by tax-loss sellers. This is consistent with the findings of Sias and Starks (1997). The authors conclude that selling losers due to tax effects at the year end is explained largely by transactions of individual investors. Despite the significantly positive coefficients for the December flag and the flag for a bull market, the WD portfolios are dispersed over time and return patterns do not differ systematically.

The coefficients for the styleboxes are reported relative to the reference category medium/blend (MV). The table shows that all growth styles large/growth LG, MG, and SG are significantly more likely to window dress than the reference style medium/blend (MB). Conversely the value styles are less likely to window dress, albeit not significantly so. Investing in growth stocks is a high risk and high expected return strategy and an investors may need to look closer at holdings information to adjust the returns for risk. If this is true, then a fund manager could be tempted to window dress as it then pays off more.

Whereas cash, the most liquid of all positions, may be an indicator of window dressing, the coefficient for the liquidity measure is not significant. This can partially be explained by the fact that the growth style classes typically contain portfolios with a lower average liquidity than the other style classes. The coefficients are negative in both specifications, suggesting that window dressing activity is less likely the higher are the invisible trading costs. Index funds indeed are less likely to be window dressing as one would expect, although the coefficient is not consistently significant.

## 4.2 Window Dressing and Turnover

The following stylized example illustrates the magnitudes we would expect from window dressing activity. Assume that a fund holds total net assets worth \$10 million. At the beginning of the quarter, the fund holds a position in stock A that accounts for 2% of total net assets. The position in stock A suffers a 25% loss over the quarter, whereas the rest of the portfolio earns 10%. Stock B, which the fund did not hold, yields a return of +25% over the same quarter. The total holding period NAV return (before subtracting any expenses) for an investor would be $0.98 \times 10\% + 0.02 \times (-25\%) = 9.30\%$.

| | NAV Return | | | | Buy-and-Hold Return | | | |
|---|---|---|---|---|---|---|---|---|
| | Beginning of Quarter | | End of Quarter | | Reported Portfolio | | Initial Portfolio | |
| | Weight | Value | Weight | Value | Weight | Value | Weight | Value |
| Stock A | 2.00% | 200,000 | 1.26% | 150,000 | | | | |
| Stock B | | | | | 1.26% | 150,000 | 1.21% | 120,000 |
| Rest | 98.00% | 9,800,000 | 98.74% | 10,780,000 | 98.74% | 10,780,000 | 98.79% | 9,800,000 |
| Total | | 10,000,000 | | 10,930,000 | | 10,930,000 | | 9,920,000 |

At the end of the quarter the value of the position in stock A dropped to \$150,000. Assume the fund manager decides to window dress by selling A and buying B just before reporting. When the fund reports a position of \$150,000 in stock B we would calculate the hypothetical return on the buy-and-hold portfolio. If the fund held the position in B over the whole quarter, then its initial investment in B must have been \$120,000 (that then increased by 25%). At the beginning of the quarter, stock B would have constituted $120,000/(120,000 + 9,800,000) = 1.21\%$ of the portfolio. The return of the buy-and-hold strategy of the reported portfolio is $0.12 \times 25\% + 0.988 \times 10\% = 10.18\%$. The observed difference between the holdings-based and NAV return is $10.18\% - 9.30\% = 0.88\%$ and dividing by 90 days per quarter we get 0.98 basis points per day. To look better by one basis point a day the fund manager needs to substitute 1.43% of daily average net assets over the quarter.[20] His turnover in percent per annum hence increases by 5.72% due to this window dressing activity.

### 4.3 What is the Effect of Liquidity Costs?

Using the Breen, Hodrick and Korajczyk (2002) liquidity measure we estimate the executions costs if the fund would continuously trade over the year. Appendix A describes the methodology to calculate the price impact per half hour as a function of dollar volume traded. So far, we have used the market capitalization weighted average of these price impact coefficients for all stocks in the fund portfolio to control for liquidity costs. The measure of invisible execution costs we use in this subsection is based on the following assumptions:

(i) The fund spreads the trading activity evenly over the entire year.

[20] Turnover is the quotient of traded assets and average (daily) net assets: $150,000/(0.5 \times 10,000,000 + 0.5 \times 10,930,000) = 1.43\%$.

(ii) The liquidity of the fund portfolio is a good approximation for the invisible execution costs of what the fund trades.

The product of annual turnover and total assets under management (as of year end) returns the annual dollar amount the fund traded. Dividing by the number of trading days and the 13 half-hour periods on a typical trading day, we get the average dollar value traded over a half hour interval, and weighting by the fraction invested in stock $i$, $w_i$, the same for an individual stock in the fund portfolio. The dollar amount traded in stock $i$ per half hour is multiplied by the price impact per half hour and aggregating over all stocks in the portfolio estimates a lower bound for transaction costs:

$$\sum_i w_i \times \underbrace{\frac{\text{Price Impact (bps per \$ Traded)}}{\text{1/2 hr Period}}}_{\text{Liquidity Costs}} \times \underbrace{\frac{\text{Turnover (\% p.a.)}\times\text{Assets in \$}}{100\times250\times13} \times w_i}_{\text{Average \$ Value Traded in Stock } i \text{ per 1/2 hr}} \tag{7}$$

where $w_i$ denotes the fraction of total net assets that is invested in stock $i$.

Figure 10 shows the distributions of these trading costs by investment style. Qualitatively, the invisible execution costs cannot explain the magnitude of the underperformance of NAVs relative to a buy-and-hold strategy we observe for window dressers. A price impact of 0.1 basis points per half hour, which appears to be a high estimate in all distributions, translates into only 1.3 basis points per trading day. Assuming that the liquidity of the portfolio is representative for the traded stocks likely overstates execution costs since funds will turn over the liquid part of their portfolio more often than illiquid stocks. Falkenstein (1996) provides evidence that funds tend to trade in liquid stocks based on the quotient of volume and shares outstanding as a proxy for liquidity. Moreover, the price impact that we use seems to be in line with the numbers that Madhavan and Cheng (1997) report for block trading in up- and down markets.

## 4.4 Persistence in Window Dressing Activity

A natural hypothesis is that managers who window dress do so repeatedly. However, a window dresser may not always have the incentive to window dress, e.g. if his portfolio has had recently strong performance. Therefore, we cannot expect a window dresser to window dress every time. Persistence is nonetheless important because it will rule out managers being classified as window dressers based on a single filing. There may be other reason why the manager traded extensively before closing out the portfolio for the quarter, e.g. unexpectedly high redemptions/inflows or a change in the composition of his benchmark. Most of these reason can be ruled out by looking at persistence. Figure 11 shows a tabulation of the number of window dressed filings per fund, which can be compared to Figure 1 which shows the distribution of total filings per fund. As can be seen, only a relatively small number of funds, about 154 or 5.5% of our sample, window dressed filings more than half their filings.[21]

To shed more light on the persistence in window dressing, we investigate to what extent a manager who window dressed his last filing also is more likely to window dress the current filing as well. Table 7 shows the cross tabulation of current and lagged window dressing indicators for the managers who had at least one window dressed filing. The top number is the observed frequency while the bracketed number is the frequency expected under the assumption of independence between current and lagged window dressing. The exact Fisher test for independence in a two way table resoundingly rejects the independence hypothesis. In particular, we see almost twice the number of occurrences of two window dressed filings in a row than what would have been expected if the occurrences were independent.

# 5 Net Changes in Holdings

For this part of the analysis, we restrict attention to the subset of consecutive fund filings which satisfy the 95% matching criterion. This reduces the sample to 11,944 reported portfolios that are at most 190 days apart. For these portfolio pairs we determine the fraction of total net assets that was shifted into (out of) recently strong (weak) performing stocks. The fractions are calculated according to equation (1) in Section 3.1, i.e. the same stock price is used for two adjacent portfolios to ensure that we capture only active net trades over the quarter and not changes in portfolio weights due to asset appreciation. Similarly, we control for trades due to asset in- and outflows.

We consider buying of winners and selling of portfolio losers separately since they do not need to coincide. A fund loading up on recently stellar performers to move them up into the top ranks does not necessarily remove overall losers from the portfolio. The fund may use uninvested cash, income from dividends, or net money inflows into the fund to increase the weight on successful securities. Vice versa, an "embarrassing" stock, i.e. a weak stock within the manager's style category, need not be replaced by a quarter winner. The comparison with funds that are identified as actively trading in strong stocks and weak stocks/losers, by inspection of net changes in portfolio holdings, reveals that these funds do not exhibit the distinct behavior of the window dressed funds identified using the shape test. As expected, the NAV returns systematically underperform the hypothetical buy-and-hold returns for funds that added strong stocks between two filings. These funds likely include funds that pursue momentum strategies as well as window dressers.

[21]Funds with less less than four filings are dropped for the calculation of the fraction of window dressed portfolios.

## 5.1 Classifying Stocks as Winners and Losers Within Style Categories

Given that stated fund objectives and investment strategies most often focus on a particular style, either a stock size class and/or value-growth orientation, we define quarter strong and weak stocks within each style category separately. This subsection describes our approach to determine the style universes. The top and bottom 10% of the stocks ranked by performance are then classified as strong and weak respectively. By focusing at the performance over the past three months before the reporting date our time horizon is also different from what is typically considered in the momentum literature, but commensurate with the reporting frequency of fund managers.

A window dresser adds recently successful securities to his portfolio and/or sells losers in his portfolio. In particular, the manager may have stocks on his book which were among the worst performers in his style category (i.e. weak stocks), in which case we call these stocks "embarrassing" stocks. For most funds, the time span between two disclosures is a quarter. However, a fund manager may not only be tempted to add the quarter's top performers, but also include recently stellar stocks in order to mimic selection skill. Similarly, a "disaster" stock over the very last weeks before reporting might well be eliminated from the portfolio. Moreover, depending on the universe of stocks the fund invests in, the relevant best or worst performing stocks differ. We determine the top and bottom 10% performing stocks over the preceding three months for each reporting date and each of the nine style universes. This involves two steps: (i) Determining the style universes, and (ii) ranking the stocks within a specific style universe and a given time horizon.

We determine the core universe of stocks that is held by funds within each of the nine style box quadrants by counting how often a stock shows up in the holdings of all funds with a particular style orientation. We do not consider market capitalization weighted holdings to prevent a few very large funds from dominating the selection of the universe. This assumes that it is a stronger indicator for a stock being held by ten small funds within the same style than by one fund with ten times the market capitalization. The final universes include the 50% most frequently held stocks. The nine stock universes are recalculated each year. A more standard approach would be to determine fundamental stock characteristics. The main problem with this approach being that only about two thirds of stocks held by mutual funds can be classified when matching book-to-market values from COMPUSTAT. With our approach some borderline cases may also fall within more than one style universes, which we consider to be realistic.

Each quarter we determine the return over the previous month, the preceding two and three months (quarter). A stock by our definition qualifies as a relatively strong stock relative to its peers if within the universe it has been a top 10% performer over a given month. The same approach is used to select the relatively weak stocks. We use the terms relatively strong and weak stocks to highlight the difference from the typical definition in the literature on momentum trading. The time horizon for momentum strategies is typically longer and stocks are not classified as winners and losers within specific investment styles.

The correlation matrix in Table 5 compares recently strong and weak stocks over the three time horizons one, two, and three months. We argued in the introduction to this section, that a fund manager who contemplates exchanging stock positions to look better at the quarter end does not necessarily focus on the performance over the entire past quarter. However, the correlations illustrate that the definition of winners for the overlapping time horizons are highly correlated, somewhat more for strong stocks. For what follows we define strong and weak stocks over the three month horizon.

## 5.2 Do Managers Add Strong and Sell Weak Stocks?

For each fund the percentage of the net asset value that has been invested in strong stocks between two reporting dates is calculated. Since we do not observe offsetting trades within the quarter these percentages represent net changes. We divide this percentage by the overall net turnover between the filings to determine a ratio of buying intensity of recently strong stocks. When the trading activity in these well performing stocks is comparable to the overall turnover of the funds we would expect a distribution around 10% – the top 10% performing stocks are defined as recently strong stocks. Panel (a) in Figure 12 shows that a substantial number of the funds (1893) in the sample have ratios greater than one. A similar comparison is provided for selling weak stocks. As we defined relatively weak stocks as the bottom performers for a given style, a fund may not have any of these "embarrassing" stocks in the portfolio. In fact, Panel (b) illustrates that 2037 reported portfolios do not contain a position in an "embarrassing" stock and on average funds have a position of 2.2% in these stocks. The distribution of the ratios measuring the percentage of relatively weak stocks in the portfolio divided by the net turnover also has a fat right tail with a mean of 1.6 and 60.4% of all funds with ratios higher than one [Panel (c)]. Funds holding less than 0.25% embarrassing stocks are excluded.

Alternatively, we consider losers within the fund's own disclosed portfolio, defined as stocks with a return that is at least one standard deviation below the mean return of the fund. The results are shown in Panels (d) and (e). On average, funds hold 8.7% portfolio losers and they tend to sell off losers more often than other securities in the portfolio until the next semi-annual filing. The mean is 1.3 and 64.5% of the portfolios with at least 0.25% losers have ratios above one.

### 5.3 Comparisons of Net Holdings Changes May Not be Able to Detect Window Dressing

Above, we analyzed changes in holdings between adjacent filings. The question is how fund managers who added strong stocks or sold losers compare to the funds we identify as window dressers based on the shape test. For this purpose we inspect the top decile of buyers of relatively strong stocks, sellers of relatively weak stocks within its style, and sellers of losers within their portfolio. All three measures of holdings changes were divided by the overall net turnover to get a measure of the relative trading intensity in these stocks. Thus, we compare the right tails in the distributions in the second column of Figure 12.

The two graphs in the top row of Figure 13 show the median of the weekly return differences for the 2,678 return window dressed portfolios from the shape test. A separate line is fitted through the medians for both 13-week periods by least squares. Panels (b) and (c) displays the return difference for the 649 portfolios that were among the top 10% buyers of recently strong stocks over the quarter preceding the reporting date (weeks -13 to -1). Panels (e) and (f) show the same diagram for the 649 top 10% sellers of recently weak stocks. The pattern for selling weak stocks is less clear. What remains is the conclusion that going forward in time the trading activity does not add any value vis-à-vis the benchmark of a buy-and-hold strategy of the reported portfolio – even before accounting for expenses. As above, the median for the filings that are not classified as window dressed [Panel (d)] exhibit no systematic deviations from zero. Table 5 displays a cross tabulation of the window dressed portfolios (based on the shape test) against buyers of strong stocks and sellers of weak stocks. As can be seen the overlap is small which is an indication of the fact that analyses of net holdings changes have difficulty picking up window dressing activity.

## 6 Conclusions

To assess the performance of a fund, an investor may analyze its past returns. However, returns on stock portfolios are noisy and it is in general very difficult to fully characterize and compare the risk-return trade-off of individual fund managers' strategies. The self-declared fund objective or investment strategy does not provide much guidance in most cases. Classifications by style orientation may change quickly over time and make peer group comparisons difficult. Therefore, investors often in part rely on information about the fund's holdings to determine whether a fund manager has superior stock picking or market timing ability.

We show evidence that, due to window dressing, the reported portfolio holdings may be a misleading indicator of the types of risks a given fund took on. We argue that simply observing net trades between reporting dates, while indicative, cannot pin down the timing of trades. Determining the timing is crucial in order to distinguish window dressing from momentum trading or other legitimate active strategies.

To address the important issue of timing, we develop a methodology to identify window dressing activity based on the differences between the observed fund NAV return and the return on a hypothetical buy-and-hold strategy of the portfolio which was subsequently reported. We find that 14.8% of filings in our sample display a pattern of returns consistent with window dressing activity and about 5.5% of funds repeatedly report portfolios that are misleading. These filings tend to be from growth funds with high turnover, managers with poor recent performance versus their peers, funds with large cash positions, and funds with high expense ratios.

Importantly, the return pattern we detect for window dressers cannot be reconciled with the manager following an ex-post successful trading strategy even before accounting for fees. We can also reject the hypotheses that the observed patterns are associated with liquidity costs.

After reporting, the pre-expense returns generated by a fund do, on average, not outperform the buy-and-hold strategy. This holds true for window dressed portfolios as well. Thus, accounting for trading costs and the decrease in tax-efficiency associated with the increased turnover, the excess trading activity associated with window dressing translates into a dead-weight loss to investors. This point has been emphasized by Jeff Molitor, principal and director of portfolio review at Vanguard Funds: "All window dressing does is create transaction costs. It takes [net asset value] out of shareholder' wallets and gives it to brokers (CNNmoney, November 13, 1998)."

Our findings suggest that requiring fund managers to report their holdings more frequently will not necessarily provide more insightful information to the investor. It could simply encourage more frequent window dressing and hence increase the dead weight loss to investors due to transaction costs. Instead, requiring funds to report the dates of their top ten portfolio acquisitions and sales over the preceding quarter would likely protect an investor more effectively from costly window dressing trades. In order to safeguard proprietary trading strategies and avoid that low-skill portfolio managers simply replicate the trading strategy of successful funds, such trades could be reported with delay.

# References


Andrews, D. (2001). Testing when a parameter is on the boundary of the maintained hypothesis, *Econometrica* **69**(3).

Breen, W. J., Hodrick, L. S. and Korajczyk, R. A. (2002). Predicting equity liquidity, *Management Science* **48**(4): 470–483.

Brown, K. C., Harlow, W. V. and Starks, L. T. (1996). Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry, *Journal of Finance* **51**(1): 85–110.

Carhart, M. M., Kaniel, R., Musto, D. K. and Reed, A. V. (2003). Leaning for the tape: Evidence of gaming behavior in equity mutual funds, *Journal of Finance* **62**(2): 661–693.

Del Guercio, D. and Tkac, P. A. (2003). Star power: The effect of morningstar ratings on mutual fund flows. Working Paper.

Dyl, E. and Maberly, E. (1992). Odd-lot transactions around the turn of the year and the January Effect, *Journal of Financial and Quantitative Analysis* **27**: 591–604.

Falkenstein, E. G. (1996). Preferences for stock characteristics as revealed by mutual fund portfolio holdings, *Journal of Finance* **51**(1): 111–136.

Haugen, R. and Lakonishok, J. (1988). *The Incredible January Effect: The Stock Market's Unresolved Mystery*, Dow Jones-Irwin, Homewood, IL.

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

Jegadeesh, N. and Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations, *Journal of Finance* **56**: 699–720.

King and Wu (1997). Locally optimal one-sided tests for multi parameter hypotheses, *Economic Reviews* **16**(2): 131–156.

Lakonishok, J., Shleifer, A., Thaler, R. and Vishny, R. (1991). Window dressing by pension fund managers, *American Economic Review* **81**(2): 227–231.

Lee, C. M. C. and Ready, M. J. (1991). Inferring trade direction form intraday data, *Journal of Finance* **46**(2): 733–747.

Madhavan, A. and Cheng, M. (1997). In search of liquidity: Block traders in the upstairs and downstairs markets, *Review of Financial Studies* **10**(1): 175–203.

Musto, D. K. (1999). Investment decisions depend on portfolio disclosures, *Journal of Finance* **54**(3): 935–952.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* **59**: 347–370.

Odders-White, E. R. (2000). On the occurence and consequences of inaccurate trade classification, *Journal of Financial Markets* **3**: 259–286.

O'Neal, E. S. (1999). Mutual fund share classes and broker incentives, *Financial Analysts Journal* **55**(5): 76–87.

Reid, B. K. and Rea, J. D. (2003). Mutual fund distribution channels and distribution costs, *Perspective, Investment Company Institute* **9**(3): 1–19.

Ritter, J. R. (1988). The buying and selling behavior of individual investors at the turn of the year, *Journal of Finance* **43**(3): 701–717.

Sadka, R. (2003). Liquidity risk and asset pricing. Working Paper, Kellogg School of Management, Northwestern University.

Sharpe, W. F. (1988). Determining a fund's effective asset mix, *Investment Management Review* pp. 59–69.

Sharpe, W. F. (1992). Asset allocation: Management style and performance measurement, *Journal of Portfolio Management* **18**(2): 7–19.

Sias, R. W. and Starks, L. T. (1997). Institutions and individuals at the turn-of-the-year, *Journal of Finance* **52**(4): 1543–1562.

# A Estimating Invisible Trading Costs

Management fees are part of the expense ratio. This does not fully describe the costs of an active strategy. In particular, invisible transactions costs are not accounted for in any mutual fund report and may differ substantially across different investment styles. Our target is to estimate the costs of buying or selling $1 million worth of a particular stock. We calculate the liquidity costs of the fund as the market-cap weighted average of the price impacts of the individual stock positions. The liquidity measure captures the expected price effect when the respective underlying fund portfolio is traded.

The basis for the calculation of liquidity costs is the TAQ database, which contains intraday trades and quotes for all securities listed on the New York Stock Exchange (NYSE), the American Stock Exchange (AMEX) and the Nasdaq National Market System and SmallCap issues (Nasdaq). We closely follow the methodology by Breen, Hodrick and Korajczyk (2002) (BHK) to determine the liquidity measure for each stock held by the mutual funds in our sample. For each stock $i$ we determine a liquidity measure $l_i$. In the first step, all trades for stock $i$ on a given trading day are associated with the prevailing bid and ask quotes. As recommended by Lee and Ready (1991) and Odders-White (2000) trades are matched with quotes that have been revised at least 5 seconds prior to the time stamp of the recorded trade. If a trade is above (below) the midpoint between bid and ask, the trade is classified as a buyer (seller) initiated trade. Within each of the thirteen half-hour periods during a trading day (from 9:30am to 4:00pm) we calculate the net volume for each stock $i$, in thousands of shares, $NVOL_i$.

We apply the following filters to the trades and quotes from TAQ:

(i) Conditions for a valid trade are a positive recorded price, no TAQ correction indicator, and no entry for special sale conditions.

(ii) Following Sadka (2003) we exclude the opening trade, whereas Lee and Ready (1991) exclude the opening trade only if it is not preceded by a quote. We also discard all trades before 9:30am.

(iii) Conditions for a valid quote are that both, bid and ask prices are positive, and the offer price is bigger or equal to the bid price.

(iv) The spread is defined as bid price minus ask price. For stocks with a quoted midpoint above $50, observations with a spread exceeding 10% of the quoted midpoint are excluded. Similarly, for stocks with a quoted midpoint of $50 and less, quotes with a spread in excess of 25% are dropped.

The slope of a time-series regression of the half-hour returns, $r_{i,\tau}$, on the half-hour net volume measures the price impact, $\beta$, in month $t$.

$$r_{i,\tau} = \alpha_{i,t} + \beta_{i,t} NVOL_{i,\tau} + \varepsilon_{i,\tau} \tag{8}$$

This time-series regression is run for each stock $i$ and each month $t$.[22] We drop the monthly observation for a ticker if we observe fewer than ten pairs of return and net turnover over the month in question, and exclude penny stocks. A stock is coined to be a penny stock whenever its minimum price during the month falls below one dollar.

As in BHK we truncate the highest and lowest beta coefficients using a 10 standard deviation bound. This eliminates on average 2 firms per month. As we would expect, only for a few outliers the price impact goes the wrong direction, i.e. is smaller the higher the net turnover. BHK normalize net volume by shares outstanding to make the coefficients comparable across firms. This allows to determine the firm characteristics of firms of different liquidity. Our goal is to assess an estimate of trading costs. Hence, we are interested in the price impact per dollar units.

The resulting sample from TAQ contains between 5,175 (April 1997) and 6,562 (March 2002) firms, with an average 5,923 firms per month. Figure 14 shows the mean, median, and the 95% confidence interval of the resulting monthly estimates over time. The coefficients exhibit a remarkable persistence over time. To avoid too much noise due to specific stock market events we take for each stock the average over the calendar year. This still allows for some degree of time variation. A successful stock might, for instance, issue new shares and become more liquid over time. The addition to a major index like the S&P 500 has also an impact on a stock's trading volume and lowers liquidity costs. The slope coefficients are then concatenated to a time series of price impact coefficients for each stock. For 47,244 out of 419,565 monthly observations (11.3%) we observe every half-hour interval within the respective month. The average of the mean and median slope coefficients are very stable over time. The mean is always above the median as the distribution is skewed towards larger price impacts.

Figure 15 exhibits the distribution of the liquidity measure for the nine style categories. As is to be expected the price impact is larger for small stocks. The price impact also increases on average if we move from value to growth stocks. This is partially explained by a larger fraction of Nasdaq stocks among growth stocks. When comparing the average price impact of the funds in Figure 15 with the mean and median price impact on NYSE, AMEX and Nasdaq stocks in Figure 14, we see that funds tend to hold the more liquid stocks on average.

# B Calculating the Pre-Expense NAV Return

The net asset value (NAV) of a mutual fund is the value of the underlying assets minus its liabilities. Dividing by the number of mutual fund shares outstanding returns the dollar value of single share, or

[22] Half-hour periods without any valid pair of trade and bid/ask quote are dropped from the sample. For a valid observation of return and net turnover at least two consecutive half-hour intervals are needed.

per share NAV. Mutual funds, legally known as open-end funds, differ from closed-end funds and Unit Investment Trusts (UIT) in that the investor directly buys the shares from the fund and not through a secondary market. Shares can be redeemed at the net asset value (NAV) on a daily basis.[23] Mutual fund shares are sold on a continuous basis unless a fund becomes too large and closes to new investors – like Fidelity Magellan in September 1997.[24]

To buy a share of the fund the investor pays the per share NAV plus any fees that the fund imposes at purchase. Depending on share class, fee structure of the fund, and the policy of the fund family, these fees are sales loads or purchase fees. When investors redeem shares of an open-end mutual fund, they sell them back to the fund. The investor receives the per share NAV minus any fees, if applicable. The fees consist of deferred sales loads and/or redemption fees.

The calculation of NAVs is regulated under the Investment Company Act of 1940 and the Rules and Regulations promulgated under that Act. Most open-end funds chose to be listed on Nasdaq and to qualify for listing, the fund is subject to the National Association of Security Dealers (NASD) rules and regulations. Under these regulations assets have to be valued at the closing prices of the major U.S. exchanges (4:00pm ET on a regular trading day) and the NAV must be reported to Nasdaq no later than 5:55pm.

### B.1 Adjusting NAVs for Dividends and Capital Gains

Distributions and expenses are accrued to date for purposes of the NAV calculation. There are two types of distributions: dividends and capital gains. Mutual funds are required by law to pay virtually all gains to their shareholders.[25] For tax purposes short-term capital gains are reported separately to investors. Short-term capital gains are realized when the fund sells an asset it has owned for twelve months or less, otherwise the capital gain distribution is classified as a long-term capital gain. Long-term capital gains are generally taxed at favorable rates, whereas short-term capital gains and dividends must be reported as ordinary income.[26] The time the fund holds a stock determines whether it is a short- or long-term capital gain, and not the length of the time period the investor holds the mutual fund share.

First, we determine the total return to the investor net of operating expenses and fees. This requires to add back all the distributions and adjust for stock splits. The daily return, adjusted for distributions and stock splits is calculated as

[23] Closed-end fund shares are typically not redeemable to the fund itself, but are traded on an exchange like a stock.

[24] Fidelity Magellan closed on September 30, 1997, to new investors outside the Fidelity retirement plans ["Trimming its sails: Magellan, the flagship of Fidelity, will close to most new investors", *The Wall Street Journal*, August 27, 1997]. At that time, the fund was the world's largest fund with $62.9 billion in assets and more than 4 million shareholders.

[25] Subchapter M of the Internal Revenue Code of 1986. Originally established under the Revenue Act of 1936.

[26] "Mutual Fund Fact Book", *ICI*, May 2003, p.19: Beginning of 2001 capital gains on assets held for more than 5 years are eligible for treatment as "qualified five-year gains" and taxed at a lower rate.

$$R^*_{t-1,t} = \frac{NAV_t + D_t}{NAV_{t-1}} SR_t - 1 \tag{9}$$

where $NAV_t$ is the net asset value at the closing of the current trading day, $NAV_{t-1}$ the net asset value at the end of the previous exchange day, $D_t$ is the sum of dividends and capital gains distributions (if any) that are reinvested at $NAV_t$. In case of share splits, the fraction $SR_t$ describes the split ratio, the number of new shares per number of old shares. Distributions are usually paid on a quarterly basis. They include dividends and capital gains net of management fees. Most funds give the investor the choice of whether the distribution should be paid out or reinvested. Reinvestments are usually exempt from any front-end sales load. Expression (9) is therefore the appropriate return under the assumption that all distributions are reinvested.

## B.2 Adding Back Operating Expenses

The expense ratio expresses the percentage of funds assets paid for operating expenses. The expense ratio accounts for management fees to the investment adviser, 12b-1 fees, and other operating expenses; such as accounting, custodial, and legal expenses, or costs of shareholder mailings. Not included are transaction costs in the form of brokerage fees. All costs for distribution services other than 12b-1 fees – front-end, deferred sales, back-end loads, maintenance fees, and redemption charges – are not part of the expense ratio since they are not paid out of the assets of the fund.[27]

Operating expenses may vary across the different fund share classes. A majority of 55% of all funds today are multi-class funds [see Reid and Rea (2003)]. Our database contains the information on the major share class with the earliest inception date – which is usually called share class A.[28] The expense ratio is recorded annually. Adjusting for the dividends and capital gains for the given share class, and adding back the operating expenses yields the same pre-expense NAV return as each share class is part of the same underlying portfolio. The holding period return from the perspective of the investor differs based on the distribution fees of the individual share classes and his investment horizon.

The total expense ratio is defined as the sum of all operating expenses over the period divided by the average daily net assets. Total expenses for an investment of $A$ are:

$$A \sum_{t=1}^{N} \left(1 - E_{daily}\right)^{t-1} \left(1 + r_{daily}\right)^{t} E_{daily}$$

The average net asset value over the $N$ days then equals

[27] These loads and fees are collected by the fund distributor, which compensates the broker or transfer agent. Depending on the share class, the investor pays the distribution costs through loads or 12b-1 fees, or a combination of the two.

[28] The Rule 18f-3 under the Investment Company Act of 1940 that allows a fund to have multiple fund share classes was adopted in 1995. The law does not specify the naming of share classes.

$$\frac{A}{N}\sum_{t=1}^{N}\left(1-E_{daily}\right)^{t}\left(1+r_{daily}\right)^{t}$$

From this we get the annual expense ratio as:

$$E_{annual}=\frac{A\times E_{daily}}{\frac{A}{N}\left(1-E_{daily}\right)}$$

Typically, annual expense ratios are in the range from 1-1.5% p.a. Solving for $E_{daily}$ and given that there are approximately 250 trading days per year

$$E_{daily}=\frac{E_{annual}}{N+E_{annual}}\approx\frac{E_{annual}}{250+E_{annual}} \tag{10}$$

We use this approximation for the daily expenses to compute the pre-expense fund return.[29]

$$R_{t}^{NAV}=\frac{1+R_{t}^{*}}{1-E_{daily}}-1 \tag{11}$$

Since beginning of 2002, the SEC requires funds to report after-tax returns for investment horizons of one, five and ten years, including and excluding costs of redemption in a standardized format.

[29] For further details see O'Neal (1999).

| | Value | Blend | Growth | Total |
|---|---|---|---|---|
| **Large** | 306 | 444 | 432 | 1,182 |
| | (568) | (796) | (646) | (2,010) |
| | 384.8 bn | 715.6 bn | 446.1 bn | 1546.5 bn |
| | (824.6 bn) | (1129.2 bn) | (702.2 bn) | (2656.0 bn) |
| **Medium** | 148 | 94 | 259 | 501 |
| | (251) | (165) | (408) | (824) |
| | 58.0 bn | 48.1 bn | 122.3 bn | 228.4 bn |
| | (110.9 bn) | (63.6 bn) | (215.8 bn) | (390.3 bn) |
| **Small** | 74 | 118 | 200 | 392 |
| | (110) | (178) | (297) | (585) |
| | 19.6 bn | 37.0 bn | 59.9 bn | 116.5 bn |
| | (26.0 bn) | (60.0 bn) | (108.0 bn) | (193.9 bn) |
| Total | 528 | 656 | 891 | 2,075 |
| | (929) | (1,139) | (1,351) | (3,419) |
| | 462.4 bn | 800.7 bn | 628.3 bn | 1891.4 bn |
| | (961.5 bn) | (1252.7 bn) | (1026.0 bn) | (3240.3 bn) |

**Table 1: Number of funds and total market capitalization in 2001.**
The number of funds and the total market capitalization in billions is tabulated for the nine Morningstar style box categories. The values are displayed for the subset of funds where we can match at least 95% of the holdings (including cash) with daily stock returns from CRSP for at least one of the fund's filings in 2001. Numbers in brackets below refer to the full sample of 4,025 U.S. domestic equity mutual funds. The total sample over the period 1997-2002 contains 3,289 (4,025) funds. Out of these, 1,214 (606) funds did not appear in our subset in 2001 or were not assigned to a style box.

| **Variable** | **Style** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | LV | LB | LG | MV | MB | MG | SV | SB | SG |
| # of Stocks | 93 | 217 | 79 | 73 | 262 | 78 | 230 | 293 | 146 |
| | (66) | (96) | (56) | (49) | (108) | (68) | (98) | (103) | (97) |
| Assets (Millions) | 1258 | 1612 | 1033 | 392 | 512 | 472 | 264 | 313 | 300 |
| | (144) | (213) | (146) | (73) | (118) | (101) | (60) | (96) | (120) |
| Price Impact (bps/$1M) | 12.00 | 9.4 | 10.8 | 38.2 | 63.8 | 56.3 | 358.2 | 220.1 | 231.1 |
| Cash | 2.3% | 1.6% | 1.7% | 2.3% | 2.8% | 1.7% | 2.2% | 2.2% | 1.6% |
| Expense Ratio | 1.2% | 1.1% | 1.4% | 1.3% | 1.2% | 1.5% | 1.4% | 1.3% | 1.7% |
| Turnover | 118.9% | 68.7% | 140.1% | 93.6% | 139.6% | 168.8% | 103.3% | 79.0% | 128.4% |
| Rating (1-5 Stars) | 2.4 | 1.8 | 1.4 | 2.5 | 2.1 | 1.6 | 2.7 | 2.1 | 2.0 |
| Index Funds | 2.9% | 18.2% | 2.8% | 4.1% | 12.8% | 1.9% | 5.4% | 9.3% | 4.0% |

**Table 2: Fund characteristics of 2,075 funds in 2001 by Morningstar style categories.**
All fund characteristics are averages over the investement style. Values in brackets indicate medians. The three size classes in the style box are abbreviated as Large, Medium, and Small; the investment orientations are Value, Blend, or Growth. The liquidity of a fund portfolio is measured as the market capitalization weighted average price impact (in basis points) if 1 million of the fund portfolio is sold within half an hour. Expense ratios and turnover are expressed as a percentage of total fund net asset values. Morningstar's rating system assigns 1-5 stars by comparing the risk-adjusted past returns within peer groups.
The total sample over the period 1997-2002, for which 95% of the holdings could be matched, contains 3,289 funds. 1,214 funds did not appear in our subset in 2001 or were not assigned to a style box.

| Model | WD (5%) | | WD (1%) | |
|---|---|---|---|---|
| | I-5 | II-5 | I-1 | II-1 |
| Number of Observations | 11834 | 11834 | 11834 | 11834 |
| Pseudo R-Squared | 6.58% | 6.69% | 8.08% | 7.55% |
| Fund Size (log10) | 0.115 ** | 0.129 ** | 0.052 | |
| # of Stocks (log10) | -0.047 | | -0.388 * | -0.426 ** |
| Cash Holdings | 0.077 ** | 0.077 ** | 0.075 ** | 0.084 ** |
| Turnover (% p.a., log10) | 0.002 ** | 0.002 ** | 0.002 ** | 0.002 ** |
| Expense Ratio | 27.917 ** | 28.078 ** | 28.014 ** | |
| Liquidity Costs | -0.032 | -0.034 * | -0.037 | |
| Performance vs. Peers | -1.366 ** | -1.381 ** | -1.958 ** | -1.888 ** |
| Momentum Trader | 0.108 | | -0.064 | |
| Indexfund | -0.681 ** | -0.707 ** | -0.344 | |
| Dead Fund | 0.407 ** | 0.341 ** | 0.464 ** | 0.399 ** |
| December Filing | 0.206 ** | 0.211 ** | 0.059 | |
| Bull Market | 0.214 ** | 0.208 ** | 0.133 | |
| Morningstar Rated | 0.094 | | 0.159 | |
| Large Value | -0.258 | -0.221 * | -0.149 | |
| Large Blend | 0.005 | | 0.208 | 0.309 * |
| Large Growth | 0.536 ** | 0.567 ** | 0.814 ** | 0.973 ** |
| Medium Value | -0.394 * | -0.331 * | -0.705 * | -0.594 * |
| Medium Growth | 0.855 ** | 0.887 ** | 1.304 ** | 1.442 ** |
| Small Value | -0.281 | | -0.117 | |
| Small Blend | -0.017 | | 0.044 | |
| Small Growth | 0.944 ** | 0.974 ** | 1.414 ** | 1.484 ** |
| Constant | -3.696 ** | -3.859 ** | -3.949 ** | -3.092 ** |

**Table 3: Fund characteristics of window dressed (WD) portfolios.**
Probit estimates of the model $P(y_i = 1|x_i) = \Phi(x_i\beta)$ where the characteristics $x_i$ are listed in the first column. The variable $y_i$ is a 0-1 variable indicating indicating whether the shape test found filing $i$ to be window dressed ($y_i = 1$). For the models I-5 and II-5 a significance level of 5% was used for the Shape test while the models I-1 and II-1 indicate that a 1% level of significance was used. The columns marked I-5, II-5, I-1 and II-1 contain the estimated coefficients $\beta$ for the probit model. A * indicates significance at the 5% level and a ** significance at the 1% level for the estimated parameter. The columns II-5 and II-1 show the results for a reduced model where insignificant variables have been removed by successive testing at the 5% significance level. Style orientation is coded as a categorical variable and the coefficients are estimated relative to the category medium blend (MB). Similarly, the coefficient for Mornigstar rated is relative to not being rated, and the flag indexfund is relative to not being an index fund. The Styleboxes are (1=LV, 2=LB, 3=LG, 4=MV, 5=MB, 6=MG, 7=SV, 8=SB, 9=SG)

| Year | Non-December | December |
|---|---|---|
| 1997 | 239 | 95 |
| 1998 | 341 | 301 |
| 1999 | 281 | 251 |
| 2000 | 230 | 194 |
| 2001 | 258 | 227 |
| 2002 | 261 | |
| Total | 1610 | 1068 |

**Table 4: Number of window dressed portfolios over time and in December.**
The 2,678 window dressed (WD) portfolios are identified using the shape test described in Section 3.2. The table counts each year the number of WD portfolios that occur in December and non-December months (mostly June filings). The data period is from January 1997 to June 2002.

| | Horizon | Strong Stocks | | | Weak Stocks | | |
|---|---|---|---|---|---|---|---|
| | Month(s) | 1 | 2 | 3 | 1 | 2 | 3 |
| Strong Stocks | 1 | 1.000 | | | | | |
| | 2 | 0.821 | 1.000 | | | | |
| | 3 | 0.765 | 0.849 | 1.000 | | | |
| Weak Stocks | 1 | 0.167 | 0.182 | 0.205 | 1.000 | | |
| | 2 | 0.152 | 0.122 | 0.158 | 0.617 | 1.000 | |
| | 3 | 0.127 | 0.117 | 0.115 | 0.615 | 0.633 | 1.000 |

**Table 5: Correlation matrix of quarter winners (losers) with well (badly) performing stocks over shorter time horizons.**
Within nine style universes we classify the top (bottom) 10% performing stocks as relatively strong (weak) stocks. The performance of stocks is ranked using three different, overlapping time horizons: the previous one month, two months, or three months. The table shows the correlations among the three alternative classifications.

| | Buyers of Strong Stocks | Sellers of Weak Stocks | Sellers of Portfolio Losers | Shape Test WDs |
|---|---|---|---|---|
| Buyers of Strong Stocks | 938 | | | |
| Sellers of Weak Stocks | 70 | 603 | | |
| Sellers of Portfolio Losers | 107 | 267 | 894 | |
| Shape Test WDs | 243 | 80 | 132 | 1,805 |

**Table 6: Freqency tabulation of buyers of strong stocks and sellers of weak stocks against the filings identified as window dressed (WD) based on the shape test.**
Using net changes between two adjacent portfolio dates (less than 190 days apart) we identify 1,195 portfolios that load up significantly on strong stocks and 721 dump weak stocks. Strong (weak) stocks are defined as the 10a style catergory over the past quarter. The shape test based on return differences between the hypothetical buy-and-hold strategy of the ex-post reported portfolio and the NAVs over the weeks leading up to the reporting date classifies 1,805 portfolios as WD. Note that the number of WD portfolios based on the shape test is now 1,805 (down from (2,678) because here we only consider the subset of 9,330 portfolio pairs for which the net portfolio changes can be computed. The number of buyers of strong stocks (1,195) reduces to 938 and the sellers of weak stocks (721) to 603 as we cross-tabulate with WDs from the shape test.

| | WD Next Period | | Total |
|---|---|---|---|
| WD Previous Period | No | Yes | |
| No | 8,011 | 1,465 | 9,476 |
| | (7,710) | (1,766) | 9,476 |
| Yes | 1,500 | 713 | 2,213 |
| | (1,801) | (412) | 2,213 |
| Total | 9,511 | 2,178 | 11,689 |

**Table 7: Table of persistence in window dressing.**
The table shows the frequency with which a window dressed filing is followed by another window dressed filing. The first number in each cell is the observed frequency. The number in brackets is the frequency we would have expected assuming independence. The exact Fisher test rejects the independence assumption, indicating that there indeed is evidence of (unconditional) persistence in window dressing by fund managers.

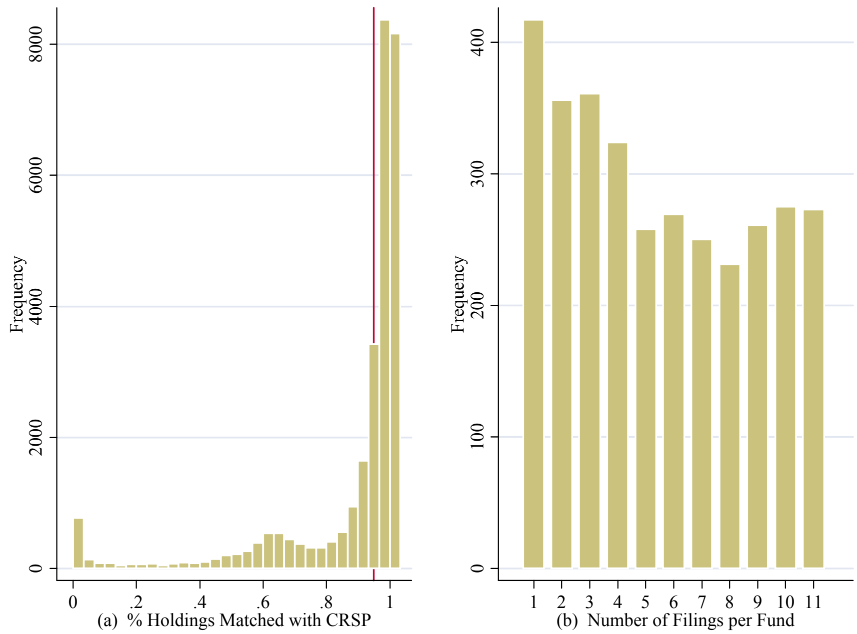


**Figure 1: Fraction of positions matched with daily stock returns for the total of 27,702 reported portfolios; and the number of filings per fund for the sample of 3,289 funds with at least 95% matching daily stock returns.**
The U.S. domestic equity mutual fund database provided by Morningstar contains 27,702 filings from 4,025 funds. Panel (a) shows the distribution of the fraction of portfolio positions that can be matched with a daily return series from CRSP. The vertical bar indicates the 95% criterion that we apply in determining the subset used to analyze window dressing activity. Panel (b) shows the number of filings per fund that survive our matching criterion. The resulting subset contains 3,289 funds and 18,139 reported portfolios.

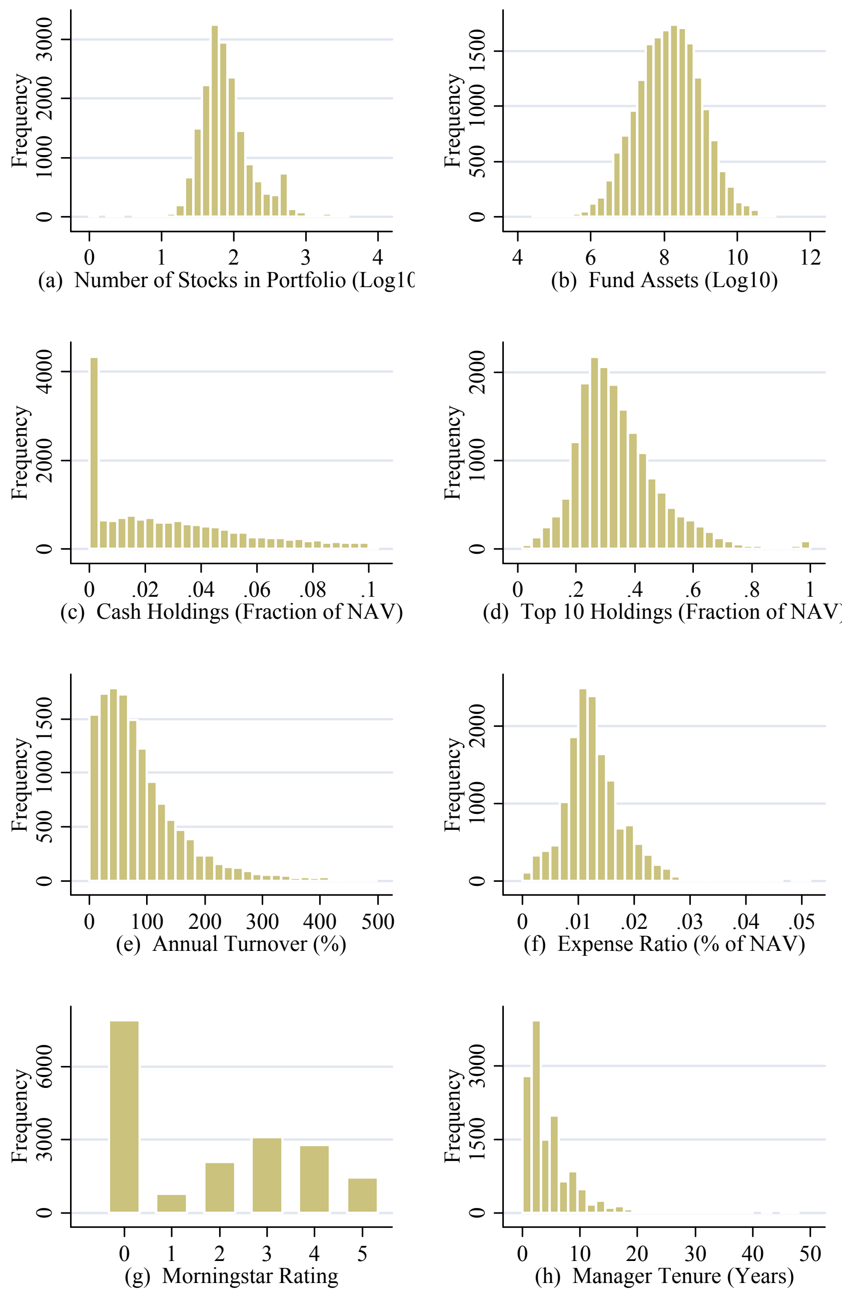


**Figure 2: Fund characteristics for the subset of 18,139 filings.**
(b) Fund size is measured as the dollar value of the net assets in millions (log10 scale). (c, d) The fund's total cash and cash equivalent position and the value of its top 10 holdings are expressed as a fraction of total net assets. (e, f) The annual turnover and the expense ratio are defined as a fraction of the average daily net asset values over the past year. The expense ratio includes management fees and asset-based compensations to the distributor and financial adviser. (g) Funds with a history of less than three years are not rated by Morningstar (rating zero). (h) Manager tenure is the number of years the manager has been in charge of a specific fund as of the filing date. The characteristics are shown for the subset of 18,139 portfolios over the six-year period 1997-2002 with 95% matching holdings.

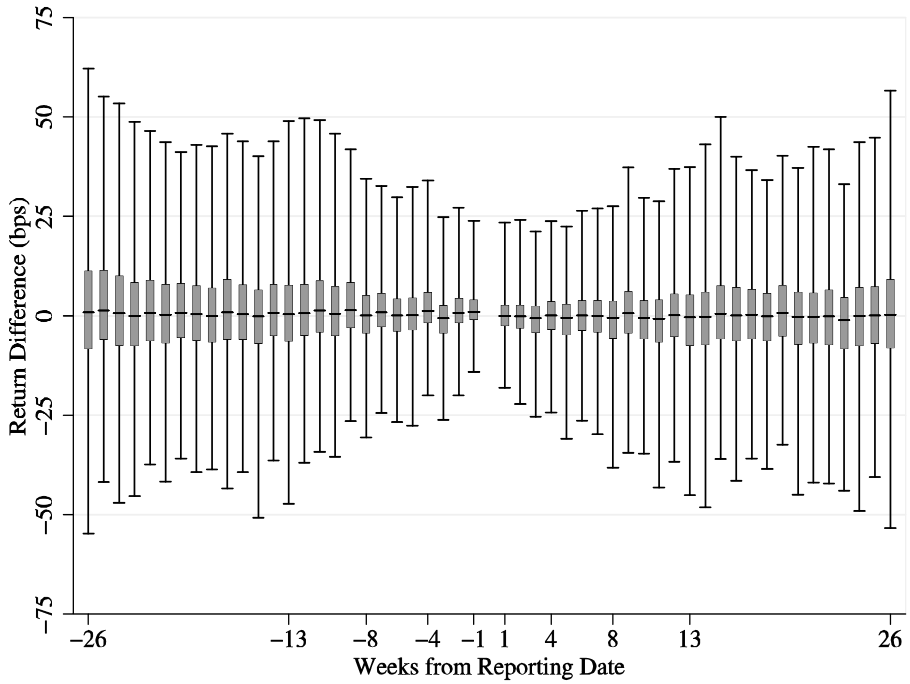


**Figure 3: Difference in returns between a hypothetical buy-and-hold strategy of the reported holdings and pre-expense net asset values (NAVs).**
For each of the 13 weeks prior to and following the reporting date we calculate the difference between the average daily returns on the reported portfolio (buy and hold) and the fund's realized pre-expense return based on reported NAVs. The difference is positive when the realized fund returns underperform the buy-and-hold strategy of the reported portfolio. The boxplot displays the distribution of these differences in basis points. The horizontal axis shows the weeks before (negative) and after (positive) the reporting date. The gray shaded boxes measure the 25th and 75th percentile, and the line drawn across the median. T-bars correspond to the 2.5 and 97.5 percentiles. The sample consists of 3,289 U.S. domestic equity mutual funds with 18,139 filings over the period from 1997 to 2002.

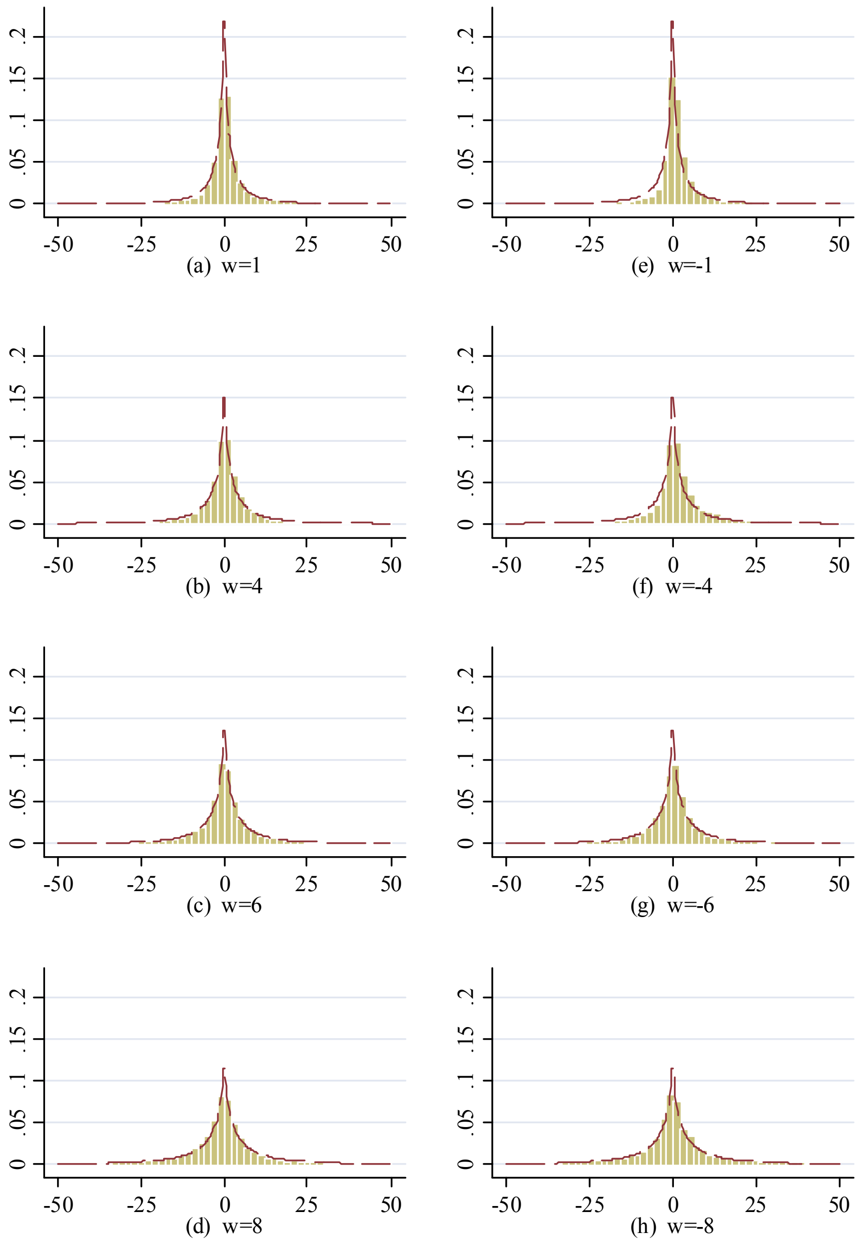


**Figure 4: Histograms of average daily return differences.**
The histograms show the average daily return difference between a buy-and-hold strategy of the reported portfolio and the NAV based return. The return differences are expressed as basis point (bps) per day. The first column shows the distributions for the weeks 1, 4, 6, and 8 following the reporting date (positive week numbers) and the second column the weeks 1, 4, 6, and 8 following the reporting date (negative week numbers). The sample covers 3,289 U.S. domestic equity funds with 18,139 filings from 1997-2002. The dashed line is the fitted GED distribution of the type described in Section 3.3. The distribution clearly fits the data post-reporting very well and reveals the skewness in the distribution of pre-reporting return differences. Note, however, that the skewness only appears during the last 4 weeks leading up to the reporting date.

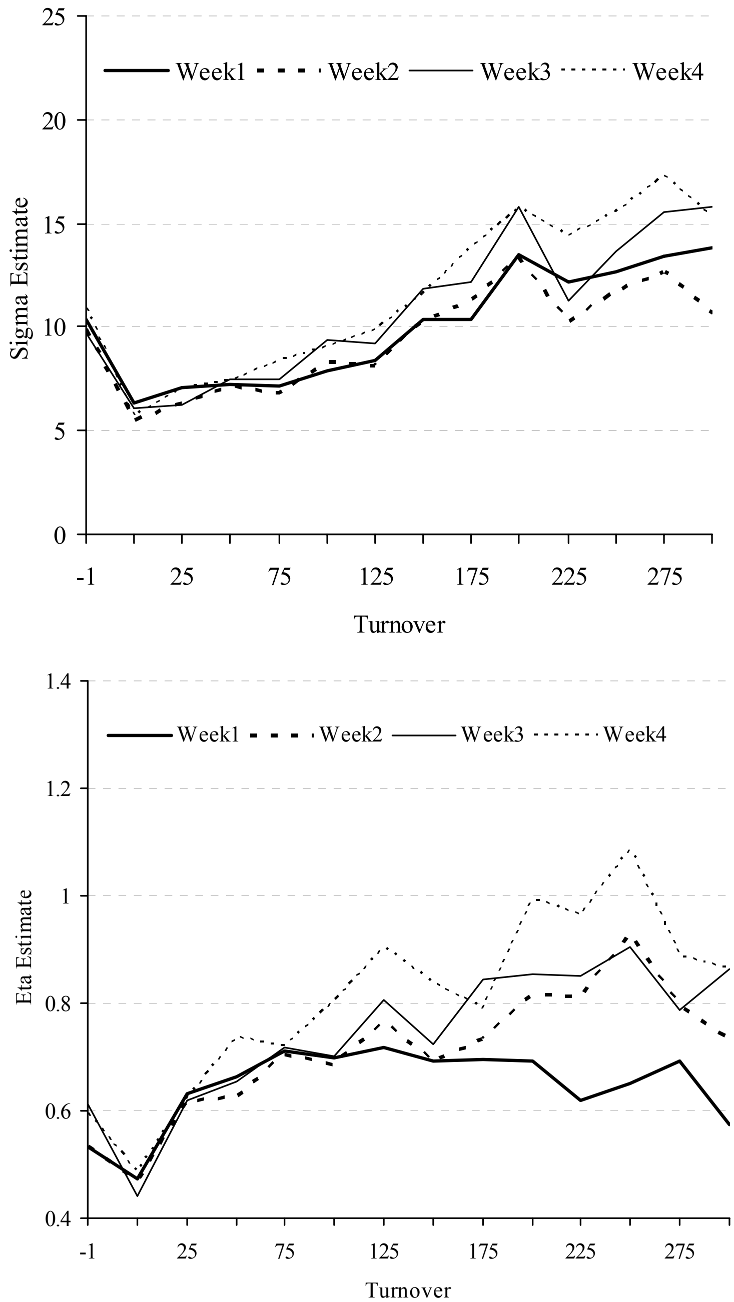


**Figure 5: Parameter estimates for the Generalized Error Distribution, by turnover category and week (post reporting).**
The average daily return differences between the reported portfolio and the fund's NAV are calculated each week prior to and after the reporting date. The return differences are assumed to be distributed according to a generalized error distribution (GED) with scale parameter $\sigma$ and shape parameter $\eta$. For the estimation of $(\sigma, \eta)$ the funds are grouped according to their turnover per annum in 25% increments (**0**: 0-25%, **25**: 25-50%, etc). The estimates are increasing in both turnover and week, corresponding to the fact that the return differences are more dispersed for higher turnover and later weeks. The category **-1** contains the fund filings for which no turnover could be assigned.

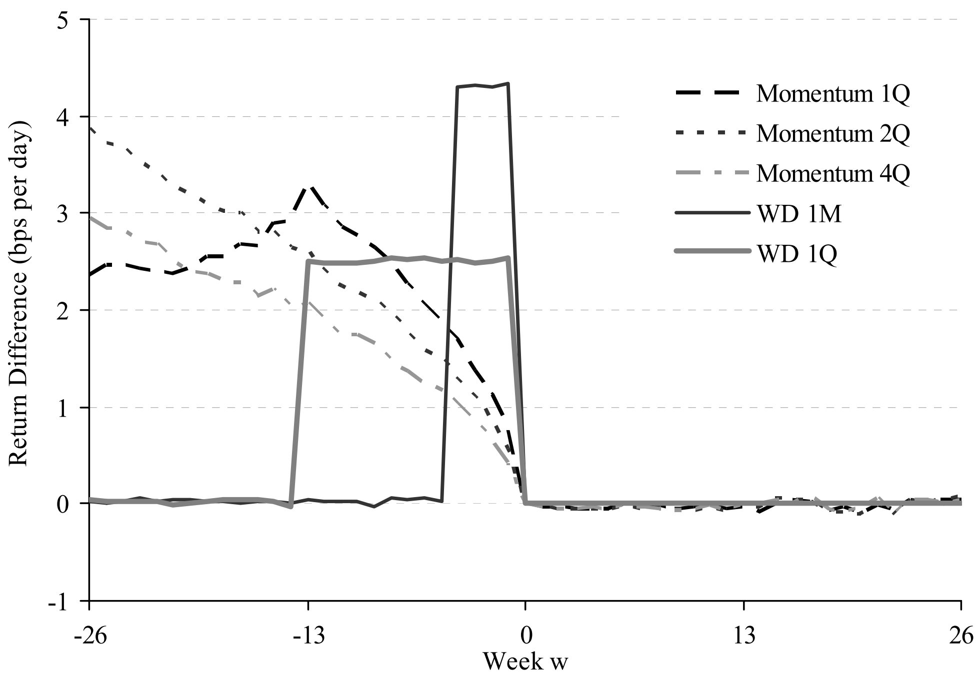


**Figure 6: Monte Carlo Simulation for a window dresser vs. a momentum trader.**
We simulate a universe of 100 stocks with returns that are serially uncorrelated and indentically distributed according to a generalized error distribution of the type estimated in Section 3.3. We choose the parameters of the GED as $\sigma = 15$, $\eta = 0.5$. The cross sectional correlation between stock returns is fixed at 0.1.
Each week a momentum trader switches 5% of his wealth from the worst performing stocks on his book into the best performing stocks over the last 1, 2, or 4 quarters. The graph shows return difference between the buy-and-hold strategy of the portfolio reported and the realized NAV returns achieved by following the momentum strategy. The difference becomes positive when the fund's NAVs underperform the reported portfolio.
For purposes of the simulation, the window dresser is assumed to be passive, except for the week of the reporting where he switches 20% of his portfolio from the worst performing stocks on his book into the best performing stocks over the last period. The horizon over which the window dresser gauges the stock performance is taken to be 4 and 13 weeks respectively.
The graphs are the average of 5000 Monte Carlo simulations.

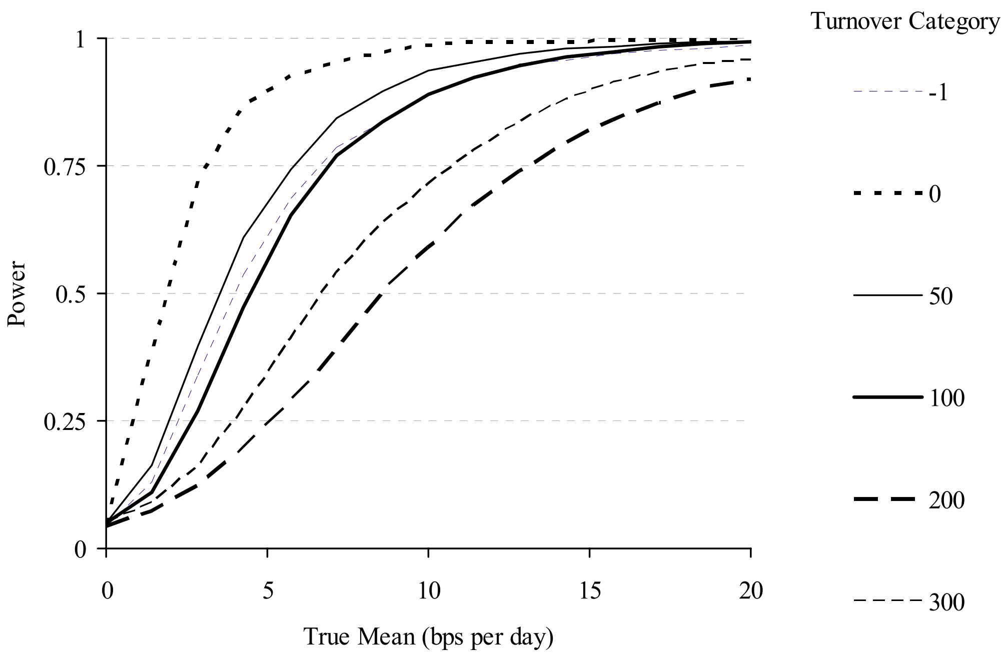


**Figure 7: The power function of the shape test by turnover category.**
The identification of potentially window dressed portfolios is based on a one sided test of $\theta = 0$ against $\theta > 0$, where $\theta$ is the average level of the return difference over the last 4 weeks of the quarter. As can be seen, higher turnover limits the ability to identify window dressing since the dispersion of the return differences is larger. The graph shows that if the true mean is $\theta = 5$ bps per day, then there is a 75% or better probability of (correctly) rejecting the null of $\theta = 0$ for the lowest turnover categories, but only a 25% probability of rejecting the null for the highest categories. The category **-1** corresponds to the funds for which no annual turnover could be assigned.

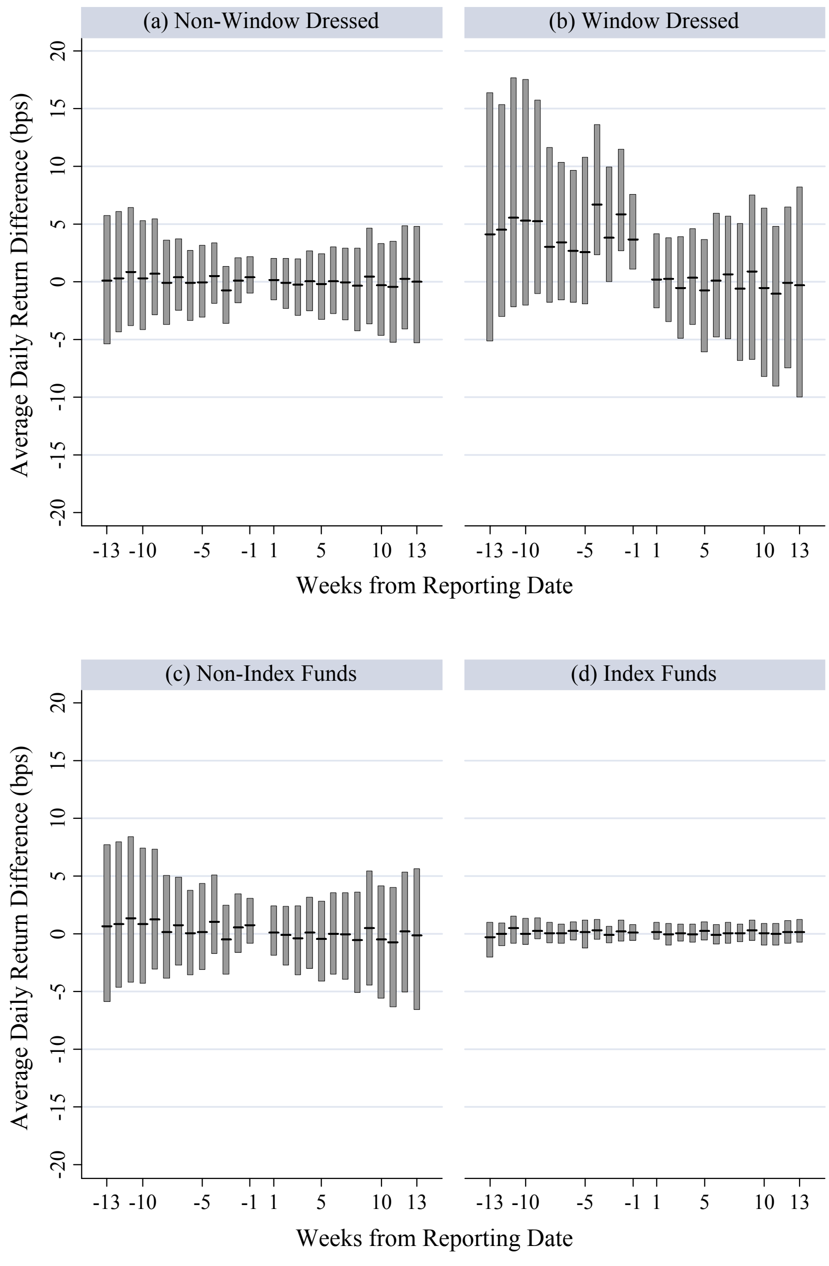


**Figure 8: Weekly return differences for window dressed and index funds.**
The graphs show the average daily return differences (returns on a buy-and-hold strategy of the reported portfolio minus the realized NAV returns) over weekly intervals in basis points per day. The results are calculated for the 13 weeks before (negative numbers) and after reporting. Thus, positive numbers indicate that the fund underperforms the reported portfolio in a given week. The gray shaded boxes show the 25th and 75th percentile, and the line drawn across the median. The first row compares the patterns for non-window dressed portfolios [Panel (a)] with the 2,678 window dressed (WD) portfolios based on the shape test [Panel (b)]. Panel (d) shows the weekly return differences for the 209 index funds with a total of 1,074 filings. Panel (c) shows the weekly return differences for the remaining 3,080 non-index funds with a total of 17,065 filings.

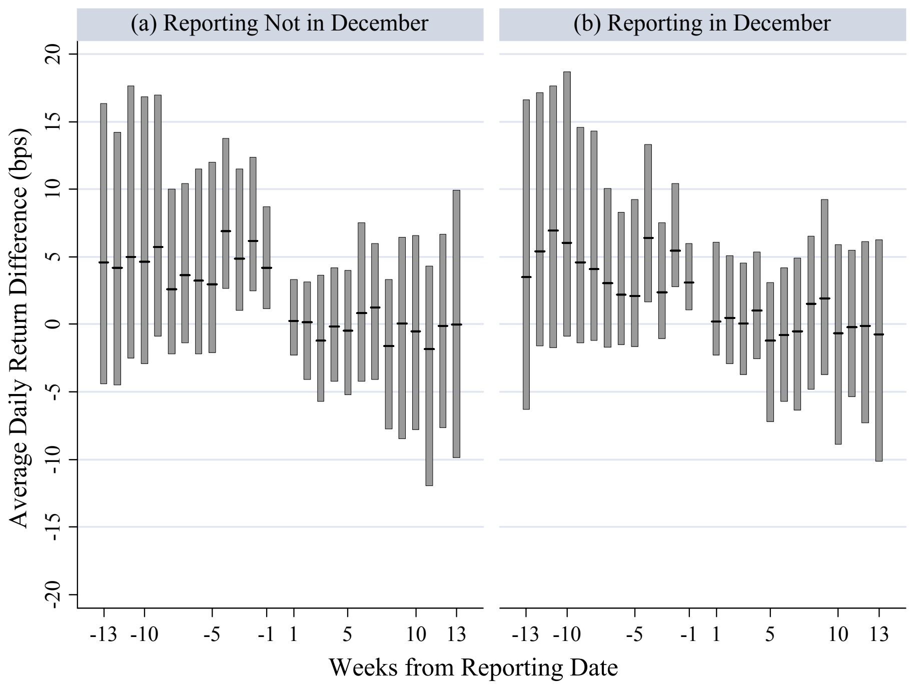


**Figure 9: Return differences for window dressed portfolios in December.**
Panel (a) Shows the subset of 1,610 window dressed portfolios with reporting dates other than December, and Panel (b) the 1,068 December filings which were identified by the shape test as likely window dressed .

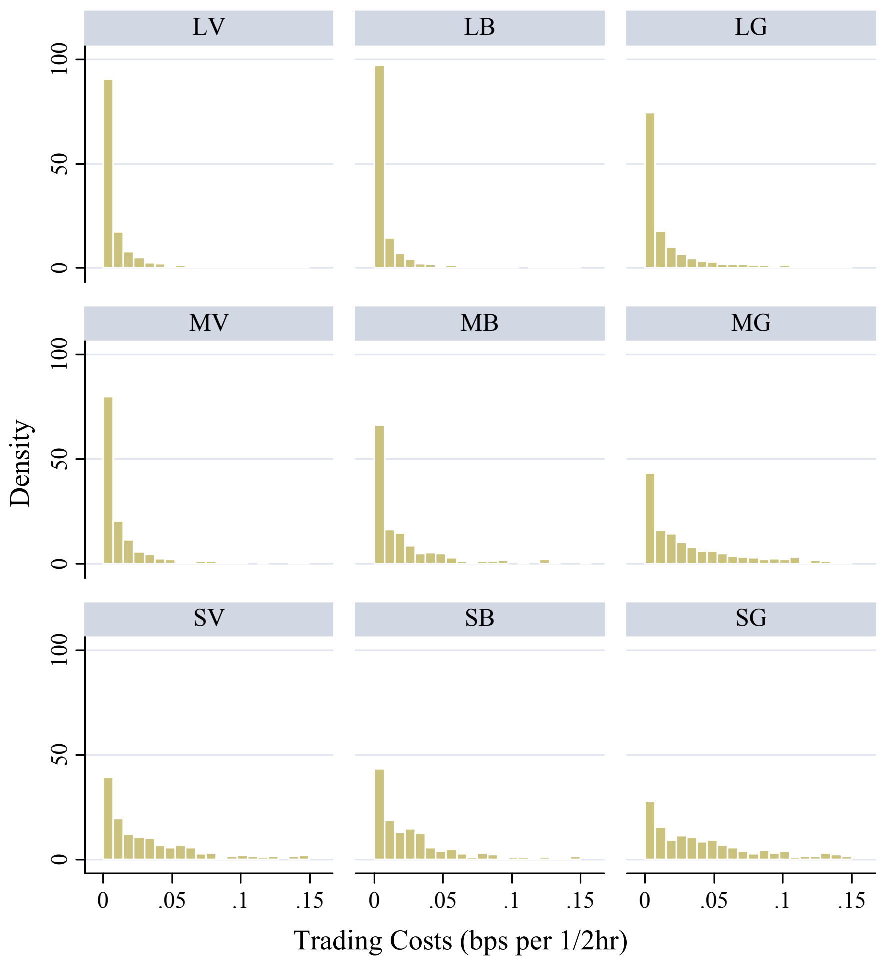


**Figure 10: Trading costs per half hour interval.**
Given the average annual turnover and the total net asset values we calculate the turnover per half hour. On a regular trading day there are 13 half-hour intervals from 9:30am to 4:00pm. Multiplying the half-hour volume by the marginal price impact per half hour returns a lower bound for the invisible transaction costs. As an estimate of the price impact we insert the BHK liquidity measure described in Appendix A. Assuming that the trading activity is evenly spread out over the entire year the half-hour turnover is:
$\sum_i w_i \frac{\text{Price Impact (bps per \$ Traded)}}{\text{1/2 hr Period}} \times \frac{\text{Turnover (\% p.a.)} \times \text{Assets in \$}}{100 \times 252 \times 13} \times w_i$
The histograms plot the price impact per half hour by investment style (Large, Medium, Small; and Value, Blend, Growth).

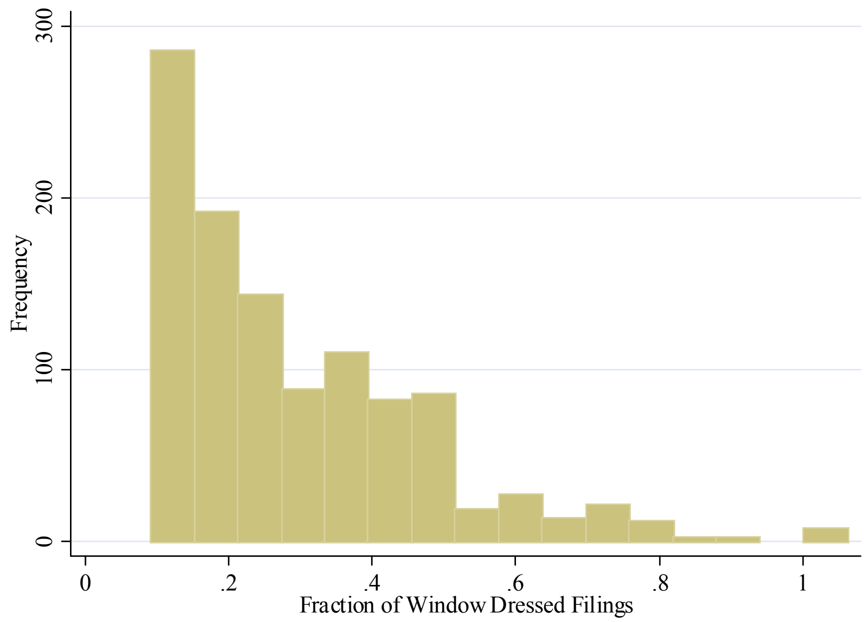


**Figure 11: Fraction of window dressed filings by fund.**
For funds that have 4 of more filings we tabulate the fraction of each fund's filings that are window dressed. We find that for about 5.5% of the funds more than half their filings are found not to be representative of the actual portfolio held by the fund during the quarter.

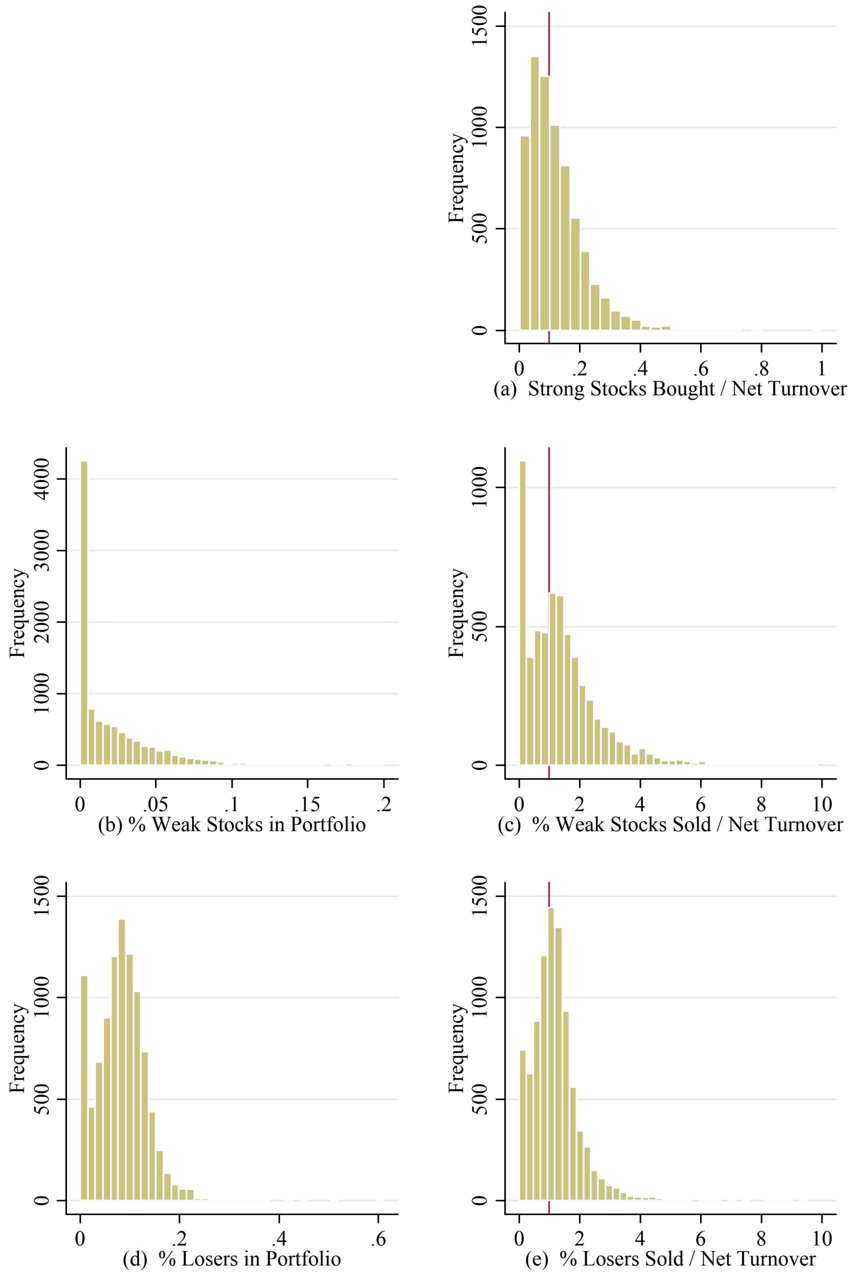


**Figure 12: Buying (selling) of recently strong (weak) stocks as a fraction of net changes in holdings.**
We classify stocks as recently strong (weak) if their cumulative returns over the last quarter are in the top (bottom) 10% within one of the nine style universes. Portfolio losers are defined as stocks with a return that is at least one standard deviation below the mean return of the fund. The histograms show the distribution of the fraction of NAV that is shifted into (out of) recently strong (weak) stocks between subsequent filings divided by the overall net turnover.

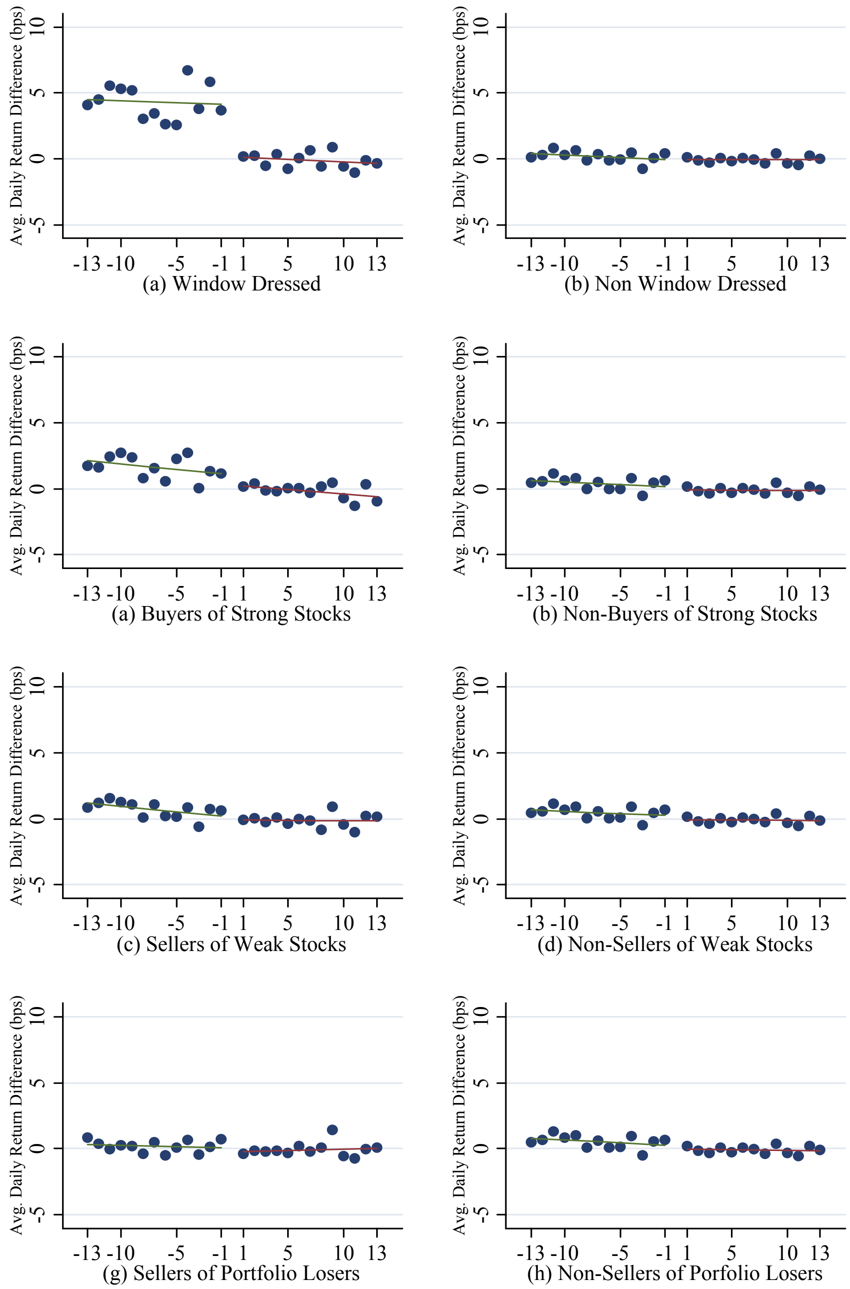


**Figure 13: Median return differences for window dressed and non-window dressed portfolios (WDs).**
Panels (a) and (b) show the median of the weekly return differences for the 2,678 return WD portfolios from the shape test. A separate line is fitted through the medians for both 13-week periods by least squares. Panels (b) and (c) displays the return difference for the 933 portfolios that were among the top 10% buyers of recently strong stocks over the quarter preceding the reporting date (weeks -13 to -1). Panels (e) and (f) show the same diagram the 933 of top 10% sellers of recently weak stocks, and (g) and (h) for the top sellers of portfolio losers. Portfolio losers are defined as stocks with a return that is at least one standard deviation below the mean return of the fund. Sample period: January 1997-June 2002.

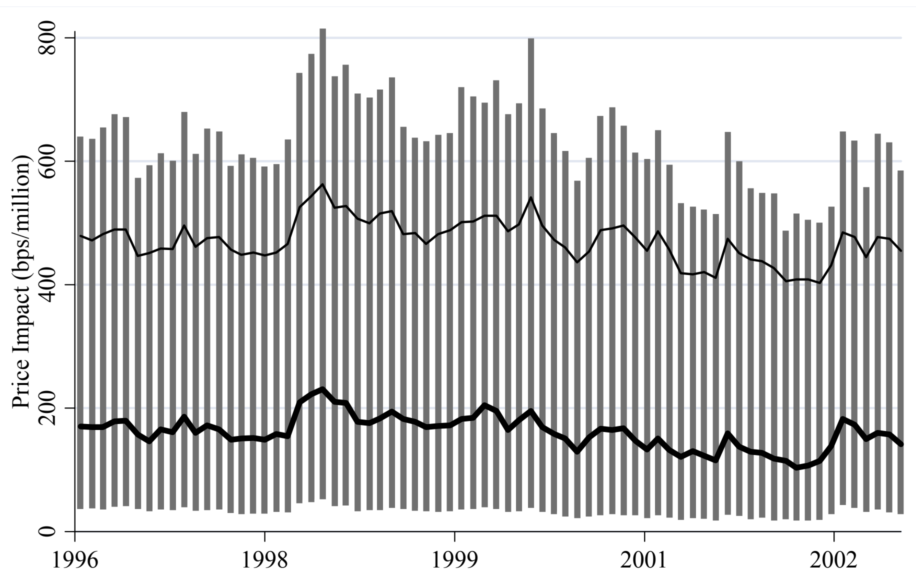


**Figure 14: Distribution of monthly price impacts for all stocks on NYSE, AMEX and Nasdaq over the period Jan. 1996 - Jun. 2002.**
The liquidity measure estimates the price impact in basis points (bps) when selling shares of the fund's underlying portfolio worth $1 million within a half-hour period. The details are described in Appendix A. The thick line shows the median and the thin line the mean. The gray bars measure the 25th and 75th percentile.

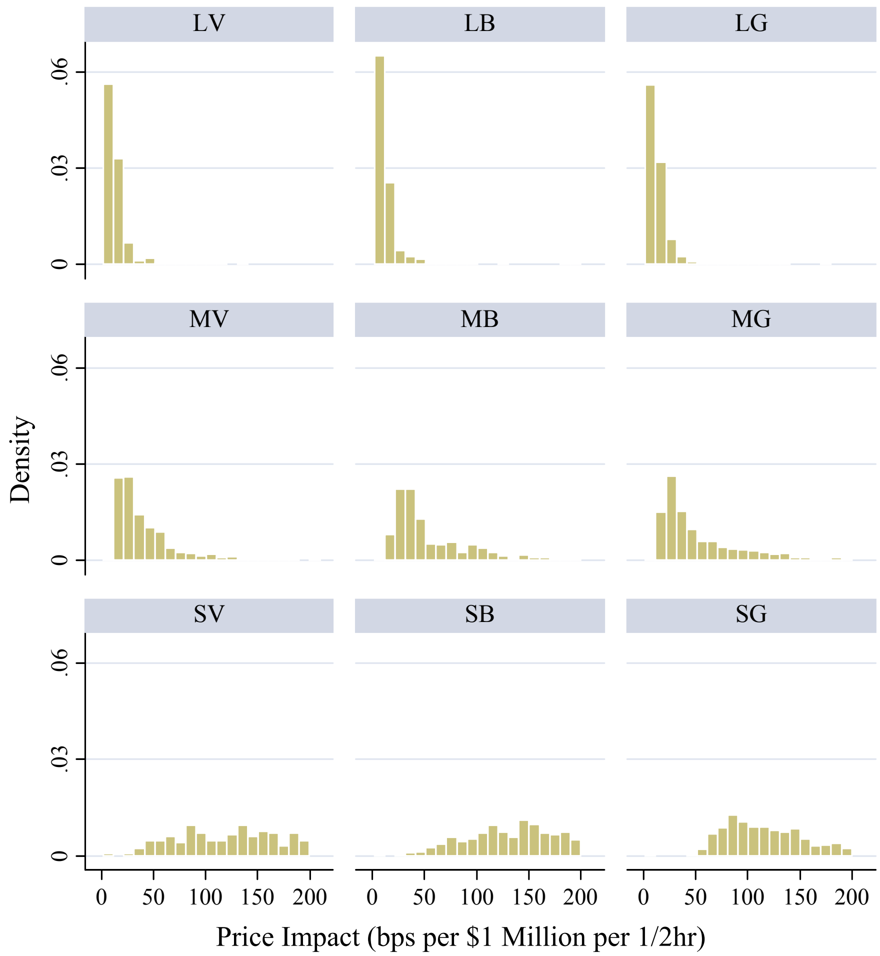


**Figure 15: Investment style and average price impact.**
Distributions of the Breen, Hodrick and Korajczyk (2002) (BHK) liquidity measure of funds within nine different investment styles. The liquidity measure estimates the price impact in basis points (bps) when selling shares of the fund's underlying portfolio worth $1 million within a half-hour period. The sample period to assess the price impact is from 1997 to 2002. The details are described in Appendix A. The rows are the three size classes (Large, Medium, and Small), the columns the style orientations (Value, Blend, Growth).