Scale effects in mutual fund performance: The role of trading costs[1]

Roger M. Edelen[2]
Echo Investment Advisors, LLC;
ReFlow Management, LLC
edelen@echoadvisors.com

Richard Evans
Carroll School of Management
Boston College
Chestnut Hill, MA 02467
evansrb@bc.edu

Gregory B. Kadlec
Pamplin College of Business
Virginia Tech
Blacksburg, VA 24060
kadlec@vt.edu

Abstract

Berk and Green (2004) argue that investment inflow at high-performing mutual funds eliminates return persistence because fund managers face diminishing returns to scale. Our study examines the role of trading costs as a source of diseconomies of scale for mutual funds. We estimate annual trading costs for a large sample of equity funds and find that they are comparable in magnitude to the expense ratio; that they have higher cross-sectional variation that is related to fund trade size; and that they have an increasingly detrimental impact on performance as the fund's relative trade size increases. Moreover, relative trade size subsumes fund size in regressions of fund returns, which suggests that trading costs are the primary source of diseconomies of scale for funds.

Current Version: March 17, 2007
First Version: September 28, 2006

JEL Codes: G11, G20, G24

[1] We thank Morningstar for providing mutual fund holdings data, John Chalmers, David Chapman, Russell Kinnel, Raman Kumar, Jeff Pontiff, Vijay Singal, and Phil Strahan for helpful suggestions. This work has also benefited from the comments of seminar participants at the 2005 Plexus Group Conference, 2006 ReFlow Symposium, 2006 Investment Management Consultants Association Conference, 2006 ICI Small Fund Conference, 2006 Mutual Fund Directors Forum, Boston College and Virginia Tech.

[2] Corresponding Author. Address: 1155 Peacock Creek, Clayton, Ca 94517. Phone: (925) 672-2667. Fax: (617) 650-878-2388. E-mail: edelen@echoadvisors.com.

Electronic copy of this paper is available at: http://ssrn.com/abstract=951367

Scale effects in mutual fund performance: The role of trading costs

Abstract

Berk and Green (2004) argue that investment inflow at high-performing mutual funds eliminates return persistence because fund managers face diminishing returns to scale. Our study examines the role of trading costs as a source of diseconomies of scale for mutual funds. We estimate annual trading costs for a large sample of equity funds and find that they are comparable in magnitude to the expense ratio; that they have higher cross-sectional variation that is related to fund trade size; and that they have an increasingly detrimental impact on performance as the fund's relative trade size increases. Moreover, relative trade size subsumes fund size in regressions of fund returns, which suggests that trading costs are the primary source of diseconomies of scale for funds.

JEL Codes: G11, G20

Electronic copy available at: http://ssrn.com/abstract=951367

I. Introduction

In an influential paper, Berk and Green (2004) argue that investment inflow at high-performing mutual funds eliminates return persistence because fund managers face diminishing returns to scale. In equilibrium, good managers do not provide investors with higher returns because the increased cost of running a larger fund offsets the managers' value added. Although Berk and Green do not explicitly model the source of diminishing returns to scale in the mutual fund industry, a likely candidate (which they point to) is trading costs. This paper explicitly examines the hypothesis that diseconomies of scale related to trading costs impede fund performance.

To test this hypothesis, we examine the relation between fund returns and trading costs for a sample of 1706 U.S. equity funds during the period 1995-2005. We use quarterly portfolio holdings data from Morningstar to infer fund trades and data from NSAR filings and NYSE TAQ to infer per-unit trading costs (brokerage commissions, spreads, and price impact) on a stock by stock basis. Annual trading costs are calculated for each fund by summing the cost of all trades during the year (i.e., the dot product of trade volume with per per-unit trade cost over all trades for the fund).

We find that annual trading costs are comparable in magnitude to the expense ratio (144 bps versus 123 bps, respectively), but have higher cross sectional variability. On average, funds fail to fully recover their trading costs -- \$1 in trading costs reduces fund assets by \$0.41. However, this average result masks substantial variation in the impact of trading costs on return performance along the dimensions of trade size and trade motive.

Consistent with the hypothesis that diseconomies of scale in fund performance are related to trading costs, we find that trading costs have an increasingly detrimental impact on performance as the fund's relative trade size increases. 'Relative trade size' refers to the fund's average trade size relative to the average trade size for all funds in the same market capitalization category (i.e., small, mid, large cap stocks).[1] The relation between trading costs and fund returns is positive for funds with a small relative trade size and negative for funds with a large relative trade size. Specifically, $1 in trading costs *increases* fund assets by roughly $0.40 for small relative trade size funds and *decreases* fund assets by roughly $0.80 for large relative trade size funds. Thus, our evidence directly establishes scale effects in trading as a source of diminishing returns to scale from active management.

Chen, Hong, Huang, and Kubik (2004) (CHHK) provide the most substantive evidence to date on the relation between fund size and performance, but they do not directly examine trading costs as a source of diminishing returns to scale.[2] They argue that trading costs alone cannot explain the observed relation between fund size and performance because a large fund could construct an internal fund of (smaller) funds, with multiple managers employing independent strategies, and thereby replicate the trading-cost efficiencies of a small fund. They propose that organizational factors, as in Stein (2002) are the source of diseconomies of scale in fund performance.

We distinguish between trading costs as a source of diseconomies of scale and other explanations (such as organizational factors) by regressing fund returns on both relative trade size and fund size. If organizational factors are responsible for the negative relation between fund performance and size, large funds should underperform small funds regardless of their

[1]A fund's relative trade size is not the same as its size. A $200 million fund that typically trades 200,000 share positions has the same relative trade size as a $2 billion fund that typically trades 200,000 share positions. From the standpoint of transaction costs, these two funds are similar in scale. Thus, relative trade size isolates diseconomies of scale related to trading from other diseconomies related to fund size.

[2] They do indirectly reference trading costs, in showing that the size effect is concentrated in small-capitalization growth funds where trading costs are likely largest. They do not, however, explicitly relate returns to trading costs (or trading activity).

relative trade size. By contrast, if diseconomies of scale in trading are responsible, large funds should under perform small funds only if they have a large relative trade size. We find that when both variables are included in the performance regressions, relative trade size supplants fund size as a cross sectional determinant of fund returns. That it, diseconomies of scale appear only to the extent that a fund actively trades in large quantities.

While our results are broadly consistent with Berk and Green (2004), they differ in one important respect. The equilibrium described in Berk and Green occurs when diseconomies of scale drive the marginal cost of operation up to the manager's marginal value added. We find that for large relative trade size funds, $1 in trading cost reduces fund value by roughly $0.80. That is, managers appear to trade well past the point where their value added exceeds the cost of transacting. We examine two potential reasons for this 'excess' trading: trading in response to investor flow, and agency costs of delegated portfolio management.

Investor flow plays a central role in the Berk and Green equilibrium: investment dollars flow to funds with positive alpha, causing increasingly high operating costs (given the fund's convex cost function). However, their model does not recognize a cost in managing the daily imbalance in inflows and outflows that most funds face. Edelen (1999) finds that funds' annual gross flow is nearly 100% of TNA and that roughly 30% of fund trades are driven by flow .[3] Because flow-induced trading is outside of the Berk and Green model (i.e., it is not related to a manager's efforts to add value) it will appear excess, or suboptimal. However, these trades are consistent with optimal fund management – despite their negative impact on return performance – because they facilitate the liquidity that fund investors demand.

[3] See for example Edelen (1999), Coval and Stafford (2005), and Alexander, Cicci, and Gibbons (2006) for analyses of flow and trading .

We use flow data (monthly gross inflows and outflows) from NSAR filings to partition trading costs into those attributable to flow and those not attributable to flow. We find that flow-induced trading partially explains the observed deviation from the Berk and Green equilibrium: Some of the value-reducing (or "excess") trading that we document is attributable to flow. However, after controlling for flow we still find evidence of excess trading by funds with large relative trade size.

We also examine the potential role of agency motives in explaining excess trading. In the context of trading, an agency cost arises when the portfolio manager acts on an incentive to execute a trade that offers less *ex ante* benefit for the investor than the cost of the trade. Agency motivated trading presumes a benefit to the advisor from the trade. A directly observable incentive for advisors to trade is soft dollar commission payments. Funds are required to report in their N-SAR filing whether the sale of fund shares; the receipt of research; or commission rebates are a consideration in choosing brokers. All three factors can give the manager an incentive to trade when the cost (to fund investors) outweighs the informational benefit of the trade, because the fund manager receives a direct benefit from the trade.

Consistent with agency-motivated trading from soft dollar relations, we find that an affirmative answer to each of the three soft-dollar questions is associated with elevated trading volume, and decreased fund returns (although the later result is not statistically significant). That is, managers appear to act on a soft-dollar induced agency motive for trading beyond the point of cost recovery. However, as with flow, controlling for soft-dollar motives we still find evidence of excess trading by funds with large relative trade size.

Dow and Gorton (1996) offer another agency explanation for fund managers' tendency to trade too much. They argue that unskilled fund managers attempt to establish a pooling equilibrium with skilled fund managers by mimicking the skilled fund managers trading activity level. Because this uninformed trading is difficult to distinguish from information-motivated trading, unskilled managers are able to justify the higher fees of skilled managers. Our evidence is consistent with a slight variation on this theory in which "skilled" and "unskilled" are considered on a *net* basis. Suppose that "gross" skill – the ability to identify valuable trade opportunities – is equally distributed across funds with large and small relative trade sizes. Then the distribution of unskilled managers on a net basis, net of trading costs, will be skewed towards funds with a large relative trade size. Funds with large relative trade sizes should therefore exhibit the most churn, as they tend to have lower net skill. This is precisely the departure from Berk and Green that we identify.

There are also arguments for excess trading based on irrational or myopic behavior. For example, Barber and Odean (2000) document biases in trading by individual investors that are difficult to reconcile with wealth maximizing objectives. Our evidence is consistent with a similar interpretation at an institutional level. However, our finding that it relates to relative trade size, and is not universally found across institutions, is a significant hurdle that both behavioral – and agency – arguments must address.

In summary, our paper makes four empirical contributions to the literature:

- Mutual funds' annual trading costs are larger in magnitude than the expense ratio. In contrast to the ambiguous relation between turnover and performance, annual trading costs bear a statistically significant negatively relation to performance.
- The negative impact of trading on performance is most pronounced for funds with a relatively large average trade size. Trading does not adversely impact performance at funds with a relatively small average trade size. Moreover, after controlling for trading

costs fund performance is no longer related to fund size. Thus, trading costs are the dominant source of diseconomies of scale in investment management.

- Flow-driven trades are shown to be significantly more costly than discretionary trades in a much larger sample and longer sample period than previously documented. This nondiscretionary trade motive partially – but not fully – explains the negative impact of trading on performance.
- Conrad, Johnson, and Wahal (2001) document that soft-dollar trades have higher costs. We show that soft dollars are also associated with substantially higher levels of trading activity and a negative impact on fund performance.

Finally, our study makes a number of methodological contributions to the literature. Many studies reference the relation between trading costs and fund performance, but there is no clear and coherent conclusion on the matter. We argue that this ambiguity stems from the use of turnover as a proxy for trading costs, and from the use of an unconditional specification of trading. Our methodology estimates funds' annual trading costs directly, on a stock-by-stock basis, rather than proxing with turnover. Our methodology also conditions on trade size and trade motive. Collectively, these enhancements provide a clearer picture on the nature of the relation between trading costs and fund performance, revealing regularities not seen with a turnover proxy.

The remainder of the paper proceeds as follows. Section II discusses methodological issues with assessing the impact of trading costs on returns. Section III details our estimates of trading costs from quarterly portfolio holdings; N-SAR filings; and NYSE TAQ transaction data. Section IV examines the impact of trading costs on fund returns, and section V examines conditional determinants of that relation – in particular, the role of fund size, flow, and agency motives for trade. Section VI concludes the study.

II. Assessing the impact of trading costs on fund return performance

Prior studies generally focus on two methodologies for evaluating the impact of mutual fund trading on returns: turnover based regressions, and comparisons of actual fund returns to hypothetical fund returns implied by portfolio holdings [e.g., Grinblatt and Titman (1989)]. This section reviews these approaches and details an alternative used in this study.

2.1. Turnover-based approach

Bogle (1994) coined the term "invisible costs" to refer to funds' expenditures on trading costs. In contrast to fund fees which are readily tabulated and widely reported through the expense ratio, trading costs are difficult to assess.[4] For this reason, much of the academic literature has used turnover as a proxy for trading costs. However, turnover is likely to be a noisy and biased proxy for trading costs.

First, turnover does not capture all of a fund's trading activity. Turnover is defined (by the SEC) as the minimum of security purchases and sales, scaled by average TNA. A fund with purchases of 100% of average TNA and sales of 20% of average TNA has lower turnover than a fund with purchases 25% of average TNA and sales of 25% of average TNA. Thus, turnover is a noisy measure of total trading activity. Moreover, it is possible that this failure to reflect trading imbalances – i.e., excess sells over buys, or vice versa imparts a bias in that, 'imbalance' trades are likely to be the result of shareholder flow. As previously noted, studies have found that flow-driven trades are more costly than discretionary trades. Thus, turnover may selectively capture relatively low cost trades, contributing ambiguity to inferences of the effects of trading costs on returns.

Second, turnover – or any other measure of trading activity – is a poor proxy for funds' trading costs because per-unit trading costs vary dramatically across assets. The Plexus Group, a leading trade-cost consultant, estimated per-unit trading costs in 2004 of 32bps (one way) for large-cap/value stocks and 132 bps for small-cap/growth stocks. Using these estimates, a large-cap value fund with turnover of 150% has lower annual trading costs than a small-cap growth fund with turnover of 50%. In such cases, turnover can yield incorrect inferences regarding the relation between trading costs and fund returns. For example, if trading is uninformed, turnover is *positively* related to fund returns whereas trading costs are negatively related to fund returns.

Finally, because turnover is reported as a percent of TNA as opposed to absolute dollar volume, it fails to capture differences in trading cost arising from differences in trade size. Both theoretical and empirical studies of market microstructure find that trade size is an important determinant of trade cost [Chan and Lakonishok (1995), Keim and Madhaven (1995)]. Thus, turnover likely understates the relative trading costs for large versus small funds, and funds with concentrated versus diffuse trading strategies.

Given these limitations, it is not surprising that the empirical evidence regarding the relation between fund returns and turnover is mixed. Elton, Gruber, Das, and Hlavka (1993) and Carhart (1997) find that fund returns are negatively related to turnover, Wermers (2000) and Chalmers, Edelen, and Kadlec (2001) find no relation between fund returns and turnover,

---

[4] Brokerage commissions are reported on the N-SAR at an aggregate level, and on the fund's Statement of Additional Information. Both reports are, however, not readily available or distributed to investors.

while Dahlquist, Engstrom and Soderlind (2000) and Chen, Jagadeesh and Wermers (2001) find that fund returns are positively related to turnover.

2.2. Hypothetical fund return approach

Grinblatt and Titman (1989) propose an indirect approach to estimating fund trading costs. In their approach, portfolio holdings data are combined with stock returns to calculate the gross return of the portfolio held by the fund for each quarter. Since this calculation does not incorporate transaction costs, the difference between this "hypothetical" fund return and the fund's actual return (before deducting expenses) is interpreted as an indirect estimate of fund trading costs. Kacperczyk, Sialm and Zheng (2006) further analyze the cross-sectional heterogeneity and persistence of this return difference which they refer to as the "return gap". They find that the return gap predicts future fund performance.

The hypothetical return approach has two shortcomings for our purposes. First, because it relies on *ex post* returns to infer trading costs, the estimate is very noisy. For example, 1/3 of the estimates yield a negative value for estimated trading cost (Kacperczyk, Sialm, and Zheng, 2006). This issue is less relevant when the analysis is restricted to large-sample averages, but inferring cross-sectional variation in trading costs and the associated impact on returns from this procedure is problematic. For this reason studies employing this measure usually employ a univariate rank-order procedure. A primary contribution of our study is the observation that the impact of trading costs on fund returns depend critically on multivariate factors.

Second, because of its indirect, residual nature, the hypothetical returns approach requires an explicit listing of (and control for) other influences on actual fund returns. Typically, this approach is implemented using only the fund's expense ratio as a control, but there can be other determinants of fund returns that do not correspond to trading costs. For example, dilution effects of stale-price arbitrageurs (Chalmers, Edelen, and Kadlec (2001), Greene and Hodges (2002)) erode actual fund performance and therefore show up as a trading cost in these measures. Likewise, a fund manager could follow a faulty investment strategy (e.g., contrarian in the face of widespread evidence on momentum (Musto and Lynch, 2003)) and the negative performance would be mistakenly construed as trading costs. Thus, while the hypothetical-returns procedure is an insightful way to characterize and measure *all* of the shortcomings associated with implementing an investment portfolio, inferences are somewhat tenuous when it comes to attributing those shortcomings to a specific factor.

2.3. A direct estimate of fund trading costs

We estimate funds' annual trading costs directly, using an approach that captures both the volume and cost per trade on a stock-by-stock basis. Specifically, we use quarterly portfolio holdings data from Morningstar to infer fund trades on a stock-by-stock basis and brokerage commission data from NSAR filings and transaction data from NYSE TAQ to infer per-unit trading costs (commissions, spreads, and price impact) on a stock by stock basis. Annual trading costs are calculated for each fund by summing the cost of all trades during the year (i.e., the dot product of trade volume with per per-unit trade cost over all trades for the fund).

While this approach to estimating trading costs has several advantages over less direct approaches, it has two limitations. First, changes in quarterly portfolio holdings do not capture all trading activity. In particular, when a fund purchases and sells the same stock within the same quarter it is not captured from changes in quarterly holdings. As we later discuss, the incidence of such slippage is about 13% at the median fund and appears to be correctable (through linear scaling). Nevertheless, it is a source of noise.

A second limitation of this approach concerns the estimation of per-unit trading cost -- particularly price impact.[5] We use the transaction data model of Hasbrouck (2004) to estimate the price-impact coefficient for each stock, each quarter. As with any statement of price impact, this approach is subject to potential misspecification of the relation between trading volume and price changes (see Chen, Stanzl and Watanabe, 2005). We attempt to address this potential shortcoming using alternative controls for price impact.

Wermers (2000) and Chalmers, Edelen, and Kadlec (2001) also use portfolio holdings data crossed with per-unit trading costs to estimate the annual trading costs for equity mutual funds. Wermers (2000) does not, however, relate trading costs to fund return performance. Rather, like many other studies, his cross-sectional analysis focuses on the relation between turnover and fund return performance. Chalmers, Edelen, and Kadlec (2001) does relate trading costs to fund return performance, however, their estimate of per-unit trading costs does not reflect price impact, the largest and most variable component of per-unit trading costs. As a result, their analysis understates the magnitude of trading costs. More importantly, the downward bias in their estimates is directly related to trade size, the primary variable of interest in our study.

Two recent studies, Chan, Faff, Gallagher and Looi (2005) and Christoffersen, Keim and Musto (2006), use fund transaction data to examine trading costs. Chan, Faff, Gallagher, and Looi, (2005) use transaction data from 26 Australian fund managers during the period 1997 to 2001. Consistent with the current study, they find that larger funds enact larger trades (relative to daily trading volume) which results in poor performance. Christoffersen, Keim, and Musto (2006) use transaction data from 210 Canadian funds during the period 2001-2004. In contrast

---

[5] Wermers (2000) uses Keim and Madhavan (1997) to interpolate per-unit trading costs based on stock and trade characteristics (stock price, trade size, market capitalization, and exchange). Because the Keim and Madhavan estimates apply to the year 1993,

to the current study and Keim and Madhaven (1997), they find that fund size is negatively related to trading costs.

III. Estimating funds' annual trading costs

This section details our estimate of annual trading costs, as applied to a large sample of open-end domestic equity funds over the period 1995-2005.

3.1. Sample selection and characteristics

Our initial sample includes 3799 open-end domestic equity funds[6] over the period 1995-2005 with quarterly portfolio holdings data available from Morningstar. We impose four constraints on this sample. First, we only have data for estimating per-unit trading costs for domestic equities, thus, we restrict our sample to funds with at least 90% domestic equity. Second, research suggests that funds with less than $15-20 million in total net assets (TNA) may have upwardly biased returns (i.e. Evans (2006)), so we exclude funds with TNA less than $20 million. Third, as in previous studies, we exclude sector funds. Fourth, we require that monthly fund returns are available for our sample from the CRSP mutual fund database.

Table 1 compares the final sample of 1,706 funds for which we have holdings data to the 1,900 open-end domestic equity funds in the CRSP mutual fund database with the constraints during this period. Our sample is slightly biased towards lower expense ratio, lower turnover funds. This may be due to the slightly larger average total net assets (TNA), but on balance, the sample is representative.

Wermers uses a historical progression of trading costs provided in Stoll (1995) to extrapolate per-unit trading costs over time.

[6] Our analysis aggregates all of the various classes of the fund into a single fund observation.

3.2. Estimating fund trading activity

We estimate fund trading activity on a stock-by-stock basis from changes in quarterly portfolio holdings adjusted for stock splits and stock mergers. To minimize the incidence of missed trades, we exclude observations where the time between reported holdings is greater than 100 days. We also exclude observations where the fund has merged with another fund, due to the difficulty in separating changes in holdings due to trading activity from changes in holdings due to the merger.

As previously noted, the primary limitation of using quarterly portfolio holdings to infer fund trades is the slippage that occurs when a stock is bought and sold between holding disclosure dates. We can, however, provide some calibration of the slippage of our estimates relative to actual fund trading activity. For 50% of our sample observations we have the total purchases and sales activity from the SEC N-SAR reports.[7] Comparing the funds' actual reported trades with that inferred from holdings data, we find that our estimates capture an average (median) of 81% (86%) of fund purchases and 82% (88%) of fund sales.

Intuitively, the degree of slippage (i.e. intra-quarter trades) is directly related to a fund's trading activity. To correct for this bias we estimate an adjustment factor by regressing actual trading activity on estimated trading activity. This regression implies a linear scaling factor of 1.22 (actual=1.22*estimated).[8] In what follows, we apply this correction to our estimates of fund trading activity. Note that, because the relation is proportional, this correction only affects the magnitude of the coefficient estimates (not t-statistics) in our cross-sectional regressions.

---

[7] The N-SAR filings are hand-matched to the CRSP database using the fund name first, and then NAV. NAV comparisons are made using the filing date of the N-SAR, and the NAVs reported in CRSP for all of the fund's share classes. This matching procedure is conservative, and almost surely removes valid data points where the NAVs differ slightly.

[8] We considered several alternative specifications for this relation including exponential, and polynomial. None of these alternatives materially improved the adjusted R-square over the linear specification.

3.3. Estimating Per-unit trading costs

For each inferred trade we apply an estimate of the per-unit cost of trading that stock. Our estimate considers three components of trading costs: brokerage commissions, effective spreads, and price impact. This section details those estimates.

3.3.1. Commissions

The fund's semi-annual N-SAR filing contains data on total brokerage commission payments, aggregated at the series level.[9] Using series-level data on trading volume, also reported in the N-SAR, we compute a commission rate for each fund for which we have N-SAR data. The commission rate is the quotient of brokerage commissions and the sum of fund purchases and sales. Panel A of Table II presents summary characteristics of the N-SAR sub-sample compared to the overall sample. Fund size, fund family size, average price of the shares traded, expense ratios and the percent of funds that charge a load are very similar between the sub-sample and the overall sample.

Because of the limited availability of commission data, we model the determinants of commissions for the matched portion of our sample, and then use the coefficients from this model to extrapolate commission rates for the non-matched funds. One of the key regressors in this model is the average price level of the stocks traded. Commissions are related to the price of the stock traded, as they are generally a fixed charge per share traded. Other regressors found to be significant relate to fund size and fees (see also Livingston and O'Neal (1996)). Commissions are negatively related to the load indicator and positively related to expense ratios. Commissions decrease with fund family size, consistent with economies of

[9] A series is a grouping of funds organized together for filing purposes, typically with no defining characteristic. For example, if one fund family acquires another, all of the target family funds may be contained in a single series.

scale or greater bargaining power on the part of fund families. By contrast, commissions are positively related to fund size.

Using the above regression coefficients, we estimate commission rates for those funds without commission data. The adjusted R-squared of the estimation regression, 18.9%, suggests a fairly high degree of noise in the extrapolation. In comparing the actual and estimated commission rates, however, we see that the median commission rate of 10.69 basis points from the N-SAR sub-sample aligns closely with the median estimated commission rate from the full sample of 10.43 basis points.

3.3.2. Effective spreads and price impact

To obtain stock-specific estimates of the effective spread and price impact we use NYSE TAQ data and the transaction data models outlined in Hasbrouck (2004). We employ a number of standard screens and filters to the TAQ data before we use it to calculate our estimates. Following Hasbrouck, we use only quotes from the primary exchange for each stock. We exclude all quotes identified as non BBO eligible (best efforts basis), or immediately following a trading halt. We exclude all trades that are identified as batched; executed as part of a basket trade; or reported out of sequence. In addition to the filters provided by TAQ, we apply filters to remove observations that may be subject to data entry errors (e.g., transposed and dropped digits). Following Keim (1989), we eliminate quotes where the bid-ask spread is greater than 20 percent of the price for stocks priced over \$10 dollars or greater than \$2 for stocks priced under \$10 dollars. We also eliminate transactions and quotes with reversals of greater than 10 percent over a sequence of three observations. To obtain a more accurate temporal ordering of trades and quotes, we adjust the time stamp for trades to correct for reporting delays in trades relative to quotes as documented in Lee and Ready (1991). However, we use a one second adjustment as opposed to a five second adjustment due to the fact that the reporting delay is substantially smaller in our more recent sample period. Finally, we eliminate all transactions that occur following a quote that was eliminated.

Following Hasbrouck (2004), we estimate the volume weighted effective spread VWS_{it} for stock i in quarter t using all transactions k during the quarter:

$$\mathrm{VWS}_{it} = \sum_{k=1}^{K}\left(\left|\frac{P_{ik} - M_{ik-}}{M_{ik-}}\right| \cdot \frac{V_{ik}}{\sum_{k=1}^{K} V_{ik}}\right)$$

where: P_{ik} is the transaction price; M_{ik-} is the midpoint of the bid and ask quotes immediately preceding transaction k; and V_{ik} is the number of shares traded.

We estimate the price impact coefficient, λ_{it}, for stock i in quarter t from the following time-series regression of changes in quote midpoints on the square root of signed trading volume using all non overlapping 15-minute time intervals, p, during the quarter:

$$\Delta M_{itp} = \lambda_{it}\sqrt{V_{itp}} + U_{itp}$$

where: ΔM_{itp} is the change in midpoint of the bid-ask quotes and V_{itp} is the signed trading volume for stock i, in quarter t, over interval p. The median estimate of λ_{it} is 0.00004 which implies that a trade of 5000 shares (median sample trade) has a price impact of 28 bp. As noted in Hasbrouck, estimates of price impact have some extreme observations that appear to be implausible. For example, the 99th percentile for our estimate of λ_{it} is 0.0007 which implies that a trade of 5000 shares has a price impact of 5%. Thus, we truncate all estimates of λ_{it} above the 95th percentile.

These price-impact estimates are based on the price-volume sensitivity of a stock as measured over a 15 minute time interval. Applying the entire quarterly holdings change to one

15 minute interval is probably not very realistic. For example, using data on actual trade packages of institutional investors, Chan and Lakonishok (1995) find that 78% of institutional trade packages are executed over two or more days. Lipson and Puckett (2006) find a 1.7 day average execution period. This suggests that the quarterly changes in our sample were likely to have been executed over a sequence of 15 minute intervals, rather than a single interval.

Unfortunately, the literature is silent on how to compound the price-impact from each period over all the periods involved with a quarterly holdings change. For this reason, we apply the entire quarterly holdings change to the price impact coefficient estimated using 15 minute intervals. This is equivalent to assuming a low accumulation rate across periods. For example, there are 26 fifteen minute intervals in a trading day. Consider a trade that is executed over one day; a sequence of 26 15-minute intervals. Our compression of the entire trade into one interval is equivalent to assuming that the price impact of trading in the last 15 minute period of the trade package is $\sqrt{26}-\sqrt{25}=10\%$ of the price impact of trading iin the first 15 minute period. While some concavity across periods is plausible, this level of concavity is perhaps too much, in which case we understate the accumulated price impact. For example, if the price impact of each period is permanent, then the compounding effect would be linear (Huberman and Stanzl, 2000). Nevertheless, any specification other than what we have assumed requires an *ad hoc* assumption as to how many periods a given position change is allocated to. To avoid this, we use the one-period assumption, recognizing that it probably understates transaction costs for large funds.

3.4. Trading cost summary statistics

Table 3 presents summary statistics of the sample funds' annual trading volume, per-unit trading costs, and annual trading costs for the overall sample and within Morningstar market-capitalization, Morningstar book-to-market, fund size, and relative trade size categories. To calculate the statistics in the table the mean (median) was calculated for each category each quarter then the overall mean was calculated from the 44 quarterly means (medians).

The first column documents annual fund trading volume (i.e., purchases plus sales) scaled by TNA. The average annual trading volume is 181% of TNA for the overall sample, and varies considerably across market capitalization and style categories. For example, the average trading volume of large capitalization stocks, 159%, is lower than that of small capitalization funds, 207%. Similarly, the average trading volume of blend funds, 144%, is lower than that of growth funds, 215%.

Columns 2-4 document the components of per-unit trading costs: percent brokerage commissions, effective spreads, and price impact. These components are summed to form the total per-unit cost in column 5. The average total one-way trade cost is 76 basis points.[10] Not surprisingly, per-unit trade costs are strongly related to the stock's market capitalization. For example, the average cost of trading small cap stocks, 146 bps is more than thee times the average cost of trading large-cap stocks -- 45 bps. However, in contrast to trading volume, there is relatively little variation in per-unit trade costs across the book-to-market categories.

Consistent with the hypothesis that trading costs cause diminishing returns to scale, Table 3 shows that per unit trading costs are about 40 bps higher for large relative trade size funds than for small relative trade size funds. By contrast, the difference is 13 bp comparing large

[10]All costs are for a single trade, i.e., not round trip.

and small TNA funds. Thus, diseconomies of scale relating to trading are best captured with the fund's average trade size, rather than the fund's size *per se* (i.e., TNA).

As noted, the primary limitation of using turnover as a proxy for trading cost is that it does not reflect differences in per-unit trading costs across asset classes. The substantial cross-sectional variation in both trading volume and per-unit trading costs documented in Table 3 suggests that accounting for both should provide sharper inferences on the impact of trading costs on fund return performance. This limitation of using turnover as a proxy for annual trading costs is best illustrated with an example using Morningstar's 3 by 3 market cap and style grid for categorizing funds. In results not tabulated, large cap growth funds have higher average trading volume than small cap value funds (184% vs. 169%), butthey have substantially lower annual trading costs (71 bps vs. 188 bps) because they have substantially lower per-unit trading costs (37 bps vs. 127 bps). This example is one of several cases where a fund category has higher average trading volume but lower average annual trading costs (i.e., annual trading volume is a *negative* proxy for annual trading costs).

Columns 6 and 7 of Table 3 document mean and median annual trading costs. For comparison purposes, columns 8 and 9 report mean and median expense ratios. Average annual trading costs, 144 bps, are comparable to average annual expense ratios, 121 bps, but the variation in annual trading costs is substantially greater than the variation in expense ratios. For example, average annual trading costs range from 77 bps (large cap funds) to 285 bps (small cap funds) whereas average expense ratios range from 112 bps (large cap funds) to 134 bps (small cap funds). Thus, annual trading costs have the potential to explain more variation in fund returns than expense ratios.

IV. The impact of trading costs on fund returns

The previous section outlines our assertion of costs incurred by each fund, each quarter, in trading portfolio securities. This section relates those cost estimates to fund performance. We use a four-factor alpha to control for factors known to relate to expected returns, as in Carhart (1996).

4.1. Trading and the momentum factor

In relating trading activity (or trading cost) to performance, it is important to control for the momentum factor. A portfolio that maintains a high loading on the momentum factor necessarily has high trading activity. This can be seen with a quintile sorting of annual trading volume (from the N-SAR filing) based on momentum loading. Observations (fund*quarters) in the highest momentum quintile have an average annual trading volume of 250% of TNA;obervations in the lowest momentum quintile have an average annual trading volume of 150% of TNA. Given the fact that momentum is also associated with positive excess returns, excluding the momentum factor from the return benchmark leads to a positive correlation between trading activity and abnormal returns. Put another way, in the absence of a direct control for the momentum effect, the relation between trading and momentum will be interpreted as a (negative) trading cost. Therefore, we use a four-factor return benchmark (Carhart (1997)) throughout our analysis.

4.2. Univariate analysis of fund performance

Figure 1 provides a bar chart of the univariate relation between fund performance and various factors. Figures 1A and 1B present the standard results in the literature, relating fund return performance to the expense ratio and total net assets. In both cases, the sorting variable appears to be negatively related to risk-adjusted performance, with a dispersion of 100 – 150 bps per year. However, the relation is not monotonic in either case. Similarly, turnover (Figure 1C) generates about 100 bps dispersion in alpha across quintiles, but again the relation is not monotonic. By contrast, trading costs (Figure D) generates about 150 bps dispersion in alpha and is monotonic. The evidence in Figure 1 suggests that trading costs sort abnormal fund returns at least as well as the other leading candidates in the literature.

4.3. Multivariate analysis of fund performance

We use the Fama-MacBeth (1973) regression framework to account for cross correlation in returns. We regress quarterly 4-factor adjusted fund returns net of expenses (Carhart (1997)) on lagged quarterly estimates of funds' trading costs, incorporating a variety of controls from the literature. Quarterly alphas are calculated as the difference between the fund's return and the product of fund's 4-factor loadings (estimated using the prior 36 months of return data) and the factor realizations. To remove the possible influences of time-series correlation in return residuals, we use the Newey-West (1987) procedure on the quarterly Fama-Macbeth (1973) coefficient estimates.

We use lagged independent variables to avoid concerns of reverse-causality. This is important for two reasons. First, flow exhibits a well documented tendency to chase past returns. Because cross sectional variation in flow generates substantial cross sectional variation in trading, the concurrent relation between trading activity (or trading costs) and return performance is positively biased (see Edelen (1999) for a detailed development and evidence on this bias). Second, many studies document a positive correlation between institutional trading in a given stock and that stock's prior return performance, consistent with trading being driven by past returns. For example, Griffin, Harris, and Topologlu (2003) argue that most of the correlation between institutional trading and returns reflects return chasing.

Similarly, Lipson and Puckett (2006) document a strong dependence of trading on lagged-returns. This lag-dependence of trading on past returns causes a positive bias between trading activity and fund returns in concurrent test, particularly at a quarterly frequency. Thus, our regressions can be regarded as both predictive and immune from this positive bias.

Table IV presents summary characteristics of the variables used in the Fama-Macbeth procedure. As in other studies of mutual funds, the number of observations increases over the sample period. Fund size, expenses, trading volume and commission rates are relatively constant over the sample period. Summing the commission, spread and price impact components of trade costs we see that per-unit trading costs decline significantly over the sample period, from 133 bps to 45 bps. Likewise, gross flows are significantly lower in the latter part of the sample than the former. Both per-unit trading costs and flow are strongly related to aggregate market volatility, so it is unclear whether this time-series variation represent secular trends or temporary distortions to a higher long-run mean.

4.3.1. Turnover and performance

Table V presents a series of Fama-Macbeth regressions. Regression 1 replicates the well-known evidence that both expenses and total net assets (TNA) negatively relate to fund performance. In regression 2, which incorporates turnover, expenses and TNA are again significant explanatory variables,. However, the coefficient on turnover is indistinguishable from zero, consistent with the literature's ambiguity on the return effects of that regressor.

4.3.2. Trading costs and performance

Regression 3 replaces turnover with the annual trade cost measure employed in this paper. Consistent with the discussion in section II, there is a significantly negative relation, -0.41 (t-statistic -1.98), between trading costs and return performance in regression 3, whereas the relation between turnover and return performance in regression 2 is insignificant. This confirms our conjecture that one must control for both trading volume and per-unit costs when assessing the impact of trading on performance. Note also that annual trading-cost coefficients are more readily interpretable than turnover coefficients, because the regressor is a direct measure of cost. For example, a coefficient greater than -1.0 implies a partial recovery of cost. Thus, our evidence suggests that on average, fund managers recover roughly half of their trading costs. As with the turnover regressions, the coefficient on TNA is significant in the absence of a control for the fund's relative trade size in regression 3. Trading costs, as estimated with the Hasbrouck square-root function, do not subsume TNA in determining fund returns.

Regression 4 expands the specification used in regression 3 by including an indicator variable equal to 1 when the average dollar value of the fund's trades is higher than the average dollar value of trades for all funds in the same market capitalization category and 0 otherwise. Comparing regression 4 with regression 3 we see that controlling for relative trade size subsumes the statistical significance of size (log TNA) in the regression (t-statistic of -1.10).

Our evidence shows that large funds do not under perform *per se*, they under perform only to the extent that they incur trading costs. The coefficient of -0.88 on annual trading cost for funds with a large relative trade size is illustrative. Trading by large relative trade size funds is highly detrimental to return performance – they recover almost none of their trading costs. By contrast, the implied coefficient on annual trade costs for small relative trade size funds is positive, but statistically insignificant– they more than recover their costs. We conclude that diminishing returns to scale in fund management can be traced to the fund's trading desk. Size *per se* is not an impediment to performance.

Separately, across the columns of Table V, regressions 1 – 4, consistent with Jensen (1968), and many subsequent studies, expenses bear a negative relation to returns.

V. Conditional analysis of trading costs and returns: Motives for trade

The coefficient on annual trading costs interacted with the large relative trade size indicator variable is -0.88, implying that a dollar of additional trade costs means $0.88 lower net returns. This result is difficult to reconcile with the the Berk and Green (2004) equilibrium, where managers choose to operate the fund in such a way that the marginal cost of operations equals the marginal value added. Our evidence indicates that managers of relatively large trade size funds trade beyond the point of cost recovery. To account for this "excess" trading we expand regression 4 by including proxies for trade motive in Regressions 5 – 8 of table V. Regression 5 includes flow, and regressions 6 – 8 include various soft-dollar dummy variables.

5.1. Flow-driven trading and fund returns

Edelen (1999) and Alexander, Cicci and Gibson (2006) provide evidence that flow-driven trades are more costly than discretionary trades. Flow driven trades are non discretionary in the sense that imbalances in shareholder flow are largely outside of the managers' control. Thus, flow might explain why funds appear to trade too much. Distinguishing between flow-driven trades and discretionary trades should provide a clearer picture of fund manager's ability to recover trading costs.

Following Edelen (1999) we obtain monthly gross inflows and outflows for approximately 50% of our sample funds from SEC N-SAR forms.[11] We estimate funds'

[11] The reduction in sample size stems from the fact that the N-SAR forms do not report the necessary ticker symbols or fund cusips that would allow us to electronically match it with our other data sources -- it must be hand-matched. After the sample is

incremental trading activity from flow by regressing funds' quarterly volume of portfolio purchases on inflows and quarterly volume of portfolio sales on outflows, while controlling for the funds cash holdings. Intuitively, large cash holdings should desensitize a funds' trading from flow to the extent that the cash is used to buffer flows. Edelen (1999) does not condition on fund cash holdings when estimating the flow-trade sensitivities. We find that the inclusion of cash holdings in the regression greatly increases the adjusted R-squared of the estimated relation. In particular the flow-trade sensitivity for inflows and outflows are: portfolio purchases = (0.5 - 0.5*cash)*inflows and portfolio sales = (0.7-0.7*cash)*outflows. These sensitivities are very close to those of Edelen (1999) who finds a flow-trade sensitivity of 0.63 for inflows and 0.76 for outflows. We use these sensitivities to estimate the volume of flow-driven trades for each fund during each quarter. We then interact this estimate of flow driven trades with our estimate of annual trade costs to assess potential differences in the cost of flow-driven trades.

Regression 5 of table V includes our estimate of flow-related trading costs along with our base-estimate of trading costs. While the flow regressor itself has a statistically significant negative coefficient (t-statistic -2.16), including flow does not materially alter inferences regarding the other variables presented in table V, with the exception of expenses (which loses statistical significance). This confirms the evidence in previous studies that flow has an adverse impact on fund returns, and provides the most direct evidence that the mechanism for this effect is trading costs. However, flow effects do not subsume the conditional dependence of the annual trading cost coefficient on relative trade size documented in the previous regressions.

---

hand-matched, we compare the NAV reported in the N-SAR to the NAV reported in CRSP to ensure the match is correct. We keep only those funds where the NAV matches exactly.

Because the Flow*TradeCost variable is an estimate of the cost of implementing trades attributable to flow, the -0.75 coefficient implies that flow-driven trades are particularly detrimental to fund performance.. After controlling for flow and scale effects, the direct effect of discretionary trading costs on fund performance is positive (though insignificant). That is, the alpha from discretionary trades of small relative trade size fund managers more than covers trading costs. By contrast, discretionary trading by large relative trade size fund managers does not, resulting in a net performance drag.

5.2. Soft-dollar related trading and fund returns

Another potential explanation for "excess" trading is soft dollars. Soft dollars refer to excess trading commissions paid to cover the cost of other goods and services received by the fund. Because soft dollars provide a mechanism to bundle trading with other services, they are a potential source of agency cost because the fund manager can benefit directly from their trades. There are a variety of non-execution related goods and services that can be paid with soft dollars including information (research), distribution (i.e., help with selling fund shares), and commission rebates (i.e., a direct reimbursement of commissions payments). Soft dollar motives are disclosed in the fund's N-SAR filing, question number 26 which asks the registrant to detail what "Considerations…affected the participation of brokers or dealers…in commissions or other compensation paid on portfolio transactions…" We use three indicator variables to indicate the fund's answer to this question – corresponding to research, distribution, and reimbursement motives for soft dollar payments.

This is a crude approach to measuring soft dollar activity. In principal, using the actual commission payments would provide a less discrete and more informative measure. Unfortunately, such data are difficult to come by. The N-SAR reports commissions, but at the series level. A series generally represents multiple funds grouped together for legal convenience, typically with no economic logic (for example, an acquisition might spawn a new series consisting of all of the funds of the target family).

Regressions 6, 7 and 8 of table V include each of the three motives for soft dollar payments, along with our base-estimate of trading costs. In each case we find a more negative coefficient on trading costs when the fund discloses a soft dollar motive. The relations are not significant, potentially due to the much smaller sample size, but their sign is more consistent with an agency cost interpretation than the alternative view that soft dollars generate useful information. However, after controlling for soft dollars, we still find that funds with a large relative trade size trade beyond the point of cost recovery.

Table VI examines the role of soft dollars in more detail, separating out the impact on per-unit trading costs (via the average commission rate) from the impact on trading volume. Regressions 1 – 3 of table VI relates commission rates to soft dollar motives, conditioning on other factors shown to be significant in Table 2. Not surprisingly, an affirmative answer to each of the three soft dollar motives correlates with a higher commission rate. Curiously, the research dummy seems to have the weakest explanatory power for commission rates (indeed, it is not statistically significant), although it is the most common affirmative answer to N-SAR question 26. Funds that disclose a reimbursement motive for soft dollars show the largest increase in commission rate (t-statistic 3.1). This may reflect a difference in the perceived cost of each disclosure: whereas commission reimbursement is relatively controversial, a research motive is a common and acceptable claim. Overall, regressions 1 – 3 show that soft dollar disclosure is associated with a higher per-unit trading cost.[12]

[12] Conrad, Johnson, and Wahal (2001) shows that soft dollars do not contribute to lower execution costs.

Regressions 4 – 6 of Table VI, relate the three soft dollar motives to the fund's trading volume. The relation is positive for all three motives for soft dollars, and the inferences are statistically significant in each case. This provides evidence that the cost of soft dollars is not simply a higher commission rate; the more important factor may be the incentive they give to fund managers to trade more. Evidently, that incentive translates into action and partially explains the "excess" trading implied by Table V.

VI. Conclusion

Berk and Green (2004) argue that inflow at high-performing mutual funds eliminates return persistence because fund managers face diminishing returns to scale. This paper examines the hypothesis that trading costs are the source of diseconomies of scale at mutual funds. Using portfolio holdings data from Morningstar and transaction-level data from NYSE TAQ, we estimate funds' annual trading costs and find them comparable in magnitude to the expense ratio (144 bps versus 121 bps). We use our estimates of annual trading costs to examine competing hypotheses regarding diseconomies of scale in fund management. We find that that trading costs supplant size as a cross-sectional determinant of fund returns. For funds with a relatively small average trade size, trading is *positively* related to fund returns. For funds with a relatively large trade size, trading is *negatively* related to fund returns. This suggests that scale effects in trading, rather than other factors in fund management, are the primary cause of diminishing returns to scale in the mutual fund industry.

While our results are generally supportive of Berk and Green (2004), there is one notable exception. We find that funds with large relative trade sizes trade well beyond the point of cost recovery, inconsistent with the premise that managers of these funds maximize value with their trading.

We examine two potential motives for this “excess” trading: investor flow and soft dollars. Both of these trade motives are outside the scope of the Berk and Green (2004) model. We find that controlling for both flow-induced trading and soft-dollar trading is important in explaining the relation between trading and performance, but neither fully explains the “excess” trading. Other possible explanations for trading beyond the point of cost recovery are the agency-cost signaling hypothesis suggested by Dow and Gorton (1996) or the possibility that managers of large funds are myopic to their trading cost handicap.

While the literature has many diverse references to trading costs and fund performance, there is not a clear and coherent conclusion on the matter. This paper provides strong evidence that should help to bring that literature together towards a coherent conclusion. On average trading costs negatively impact fund performance, but the negative impact is confined to three clearly identified factors: scale diseconomies; operational trades (i.e., flow); and agency related trades (i.e. soft dollars).

References

Alexander, G., Cicci, G., and S. Gibson, 2006, "Does motivation matter when assessing trade performance? An analysis of mutual fund trades, forthcoming *Review of Financial Studies*.

Berk, J. and R. Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy*, 112, 6, 1269-1295.

Bogle, John C., Bogle on Mutual Funds: New Perspectives for the Intelligent Investor, Irwin Professional Publishing, 1994, Burr Ridge, IL.

Carhart, M, 1997, On persistence in mutual fund performance, *Journal of Finance*, 52, 1, 57-82.

Chalmers, J., Edelen, R., and G. Kadlec, 2001, Mutual fund returns and trading costs: evidence on the value of active fund management, University of Pennsylvania working paper.

Chan, H., R.W. Faff, D.R. Gallagher, and A. Looi, 2005, "Fund Size, Fund Flow, Transaction Costs and Performance: Size Matters!", University of New South Wales working paper.

Chan, K.C. and J. Lakonishok, 1993. Institutional trades and intraday stock price behavior, *Journal of Financial Economics*, 33. 173-199.

Chan, K.C., and J. Lakonishok, 1995. The behavior of stock prices around institutional trades, *Journal of Finance*, 50, 1147-1174.

Chen, H., Noronha, and V. Singal, 2006, "Index changes and losses to index fund investors*",* *Financial Analyst Journal*, July/August, 31-62.

Chen, J., Hong, H., Huang, M., and J. Kubik, 2004, Does fund size erode mutual fund performance? The role of liquidity and organization, *American Economic Review*, December, 1276-1302.

Christoffersen, S., D. Keim, and D. Musto, Valuable information and costly liquidity: Evidence from individual mutual fund trades, University of Pennsylvania working paper.

Collins, S. and P. Mack, 1997, The optimal amount of assets under management in the mutual fund industry, *Financial Analysts Journal*, September/October, 67-73.

Conrad J., Johnson K., and S. Wahal, 2001, Institutional Trading and Soft Dollars, *Journal of Finance* 56, 397-416.

Coval J. and E. Stafford, 2005, "Asset fire sales (and purchases) in equity markets, Harvard University working paper.

Dow J., G. Gorton, 1997, Noise trading, delegated portfolio management and economic welfare *Journal of Political Economy*, October Vol 105:5 p 1024-1050.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *Journal of Finance*, 52, 1035-1058.

Edelen, R., 1999, Investor flows and the assessed performance of open-end mutual funds, *Journal of Financial Economics*, 53, 439-466.

Edelen, R., R. Evans and G. Kadlec, 2007, Expense shifting with Soft Dollars, Boston College working paper.

Edelen, R. and G. Kadlec, 2006, Agency costs of institutional trading, Virginia Tech working paper.

Elton, E., Gruber, M., Das, S., and M. Hlavka, 1993, Efficiency with costly information: a reinterpretation of evidence from managed portfolios, *Review of Financial Studies*, 6,1,1-22.

Evans, R., 2007, The Incubation Bias, Boston College working paper.

Fama, E. and J. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607-636.

Gompers, P. and A. Metrick, 1998, How are large institutions different from other investors? Why do these differences matter for equity prices and returns? Harvard University and National Bureau of Economic Research working paper.

Griffin J., Harris J., and S. Topaloglu, 2003, The dynamics of institutional and individual trading, *Journal of Finance* 58, 1515-1547.

Grinblatt, M. and S. Titman, 1989, Mutual fund performance: An analysis of quarterly holdings, *Journal of Business* 62, 393-416.

Hasbrouck, J., 2004, Trading costs and returns for U.S. equities: Evidence using daily data, Stern School of Business working paper.

Jensen, M., 1968, The performance of mutual funds in the period 1945-64, *Journal of Finance*, 23, 389-416.

Jones, C. and M. Lipson, 1999, Execution costs of institutional equity orders, *Journal of Financial Intermediation*, 8,123-140.

Kacperczyk. M., C. Sialm, and L. Zheng, 2006, “Unobserved actions of mutual funds,” Forthcoming, *Review of Financial Studies*.

Kavajecz K., and D. Keim, 2005, Packaging liquidity: blind auctions and transaction efficiencies, *Journal of Financial and Quantitative Analysis* 40, 465-492.

Keim, D. and A. Madhavan, 1995. Anatomy of the trading process, Empirical evidence on the behavior of institutional traders, *Journal of Financial Economics*. 37, 371-398.

Keim, D. and A. Madhavan, 1997. Transactions costs and investment style: an inter-exchange analysis of institutional equity trades, *Journal of Financial Economics*, 46, 265-292.

Kinnel, R., 2004, The advantages of small funds, Morningstar white paper.

Lipson, M. and A. Puckett, 2006, "Volatile markets and institutional trading", working paper, Darden Graduate School of Business.

Livingston, Miles and Edward S. O'Neal, 1996, Mutual Fund Brokerage Commissions, The Journal of Financial Research, 19, 273-292.

Lynch, A. and D. Musto, 2003, How Investors Interpret Past Fund Returns, *Journal of Finance* 58, 2033-2058.

Newey, W. and K. West, 1987, A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, 55, 703-708.

Stein, J., 2002, Information production and capital allocation: decentralized versus hierarchal firms, *Journal of Finance* 57, 5, 891-922.

Wermers, R., 2000, Mutual Fund Performance: an empirical decomposition into stock-picking talent, style, transactions costs, and expenses, *Journal of Finance*, 55, 4, 1655-1695.

Table I. – Sample Description

This table provides a comparison between the Morningstar mutual fund sample used in this paper and a sample of CRSP mutual funds with the same filters applied (share classes aggregated). Specifically both samples start with all domestic equity funds from 1995-2005 and exclude: sector funds, funds with TNA less than $20 million, and funds with average equity holdings less than 90% of their assets. Panel A reports the total number of funds, average expense ratio, annual return, annual 4-factor alpha (calculated using the factor coefficients estimates from a regression of monthly returns over the previous 36 months), total net assets (millions of dollars) and annual turnover. Panel B divides the CRSP sample into thirds by the fund's average 36 month Fama-French book-to-market (beta HML) and size (beta SMB) coefficients. Using the tercile breakpoints for the book-to-market and size coefficients from the CRSP sample, the number of funds in the Trading sample with coefficients in the same coefficient tercile is shown.

Panel A - Descriptive Statistics	CRSP	Morningstar	Overlap
Number of Funds	1900	1706	90%
Expenses	1.28%	1.22%	
Returns	9.76%	10.81%	
4-factor Alpha	-1.88%	-1.96%	
Total Net Assets ($mm)	$1,424	$ 1,448	
SEC Turnover	94%	86%	
Panel B - Number of funds, by size tercile and book-to-market tercile			
Small Cap (High beta $_{SMB}$)	633	600	95%
Mid Cap (Middle beta $_{SMB}$)	634	554	87%
Large Cap (Low beta $_{SMB}$)	633	552	87%
Value (High beta $_{HML}$)	633	560	88%
Blend (Middle beta $_{HML}$)	634	570	90%
Growth (Low beta $_{HML}$)	633	576	91%

Table II. – Commission Rates

A model for commissions is estimated using data from the SEC form N-SAR. N-SAR data from 1995 to 2005 is collected and hand-matched with the CRSP mutual fund database. Bond, balanced, international and precious metal funds are removed. Net asset values (NAV) from the N-SARs are then compared to NAVs reported in CRSP. Only those funds that have an exact match are retained. The "Commissions" sample contains these matched funds. As a comparison, the descriptive statistics from the full trading sample used in the paper are also provided. The commissions data contains observations from two types of N-SAR structure; multi and single-fund trusts. Single-fund trusts include only one fund per filing. Multi-fund trusts include multiple funds per filing. Because the brokerage commission data in form N-SAR is only reported in aggregate by trust, the trust average is assumed to be correct for all funds in the trust in the multi-fund trusts. The dependent variable in these regressions is the commission ratio which is calculated as the trust-level brokerage commissions (N-SAR item 21) divided by the sum of the aggregate trust-level purchases and sales (N-SAR items 71 A and B). The independent variables include an intercept, the fund family size quintile (1=smallest, 5=largest), the natural log of the fund TNA (in millions), the average trade share price (in $), the expense ratio (in %) and an id variable indicating whether or not the fund charges a load (load = 1, no-load = 0). Standard errors are calculated clustering by fund. The asterisks denote statistical significance as follows: *** - significant 0.1%, ** - significant at 1%, and * - significant at 5%. Standard errors are included in parentheses.

Panel A. Data	Trust Structure	
	Commission Sample	Full Trading Sample
Mean Fund Size ($millions)	$1,235	$1,448
Mean Fund Family Size ($billions)	$246.9	$271.5
Mean Price of Shares Traded	$31.56	$31.55
Mean Expense Ratio	1.18%	1.22%
Fraction of Load Funds	44%	49%
Fund-Quarter Observations	5,815	23,370

Panel B. Regression model		
Intercept		8.62***
Fund Family Size Quintile		-0.68***
Log(Fund Size)		0.61***
Load ID (=1 if Load Fund)		-1.66***
Average Shareprice Traded		-0.14***
Expense Ratio		4.88***
Observations		5,815
R-Squared		18.92%
Median Commissions (bp)	Subsample	10.69
Median Estimated Commissions (bp)	Full Sample	10.43

Table III – Annual Trading Costs

This table reports descriptive statistics of sample funds' annual trading volume, per-unit trade costs, and annual trading costs. To calculate the statistics in the table, first the mean (median) was calculated for each category in each quarter. Then, the mean (median) reported in the table is calculated from this distribution of 44 quarterly means (medians). The first column reports annual trading volume (total purchases and sales of equity securities calculated from quarterly holdings data) scaled by TNA. The next four columns report three components of per-unit trading costs: brokerage commissions, bid-ask spreads, and price impact (estimated on a stock-by-stock basis using N-SAR commisions data and TAQ transactions data), and total per-unit trading costs each presented as a fraction of dollars traded. The sixth column reports annual fund trading costs calculated as the dot product of all trades made during the quareter with the per-unit cost of each trade, scaled by TNA. The eighth and ninth columns report statistics about the expense ratio. The table reports means for the entire sample (All) and then repeats the analysis for four separate splits of the sample: by relative trade size (Large Relative Trade Size and Small Relative Trade Size), by the market-capitalization of stock (Large-cap Stks, Mid-cap Stks, Small-cap Stks) traded by the fund, by the book-to-market of the stocks (Value, Blend, Growth) traded by the fund and by the size of the fund itself (Large TNA and Small TNA). The market-capitalization and book-to-market categories for each fund are given in the holdings data and are the same designations as used by Morningstar. The Large TNA and Small TNA categories represent a split of the sample by median fund size.

	Trading Volume	Per-Unit Trading Costs				Annual Trading Costs Volume*Per-Unit Costs		Expense Ratio	
Fund Group:	Mean	Commissions	Mean + Spread +	Price Impact=	**Per-Unit Cost**	Mean	Median	Mean	Median
All	**181%**	0.13%	0.13%	0.49%	**0.76%**	**1.44%**	0.89%	1.21%	1.17%
Large Relative Trade Size	**195%**	0.12%	0.13%	0.68%	**0.94%**	**1.89%**	1.28%	1.21%	1.15%
Small Relative Trade Size	**165%**	0.13%	0.13%	0.30%	**0.56%**	**0.96%**	0.57%	1.20%	1.20%
Large-cap Stks	**159%**	0.11%	0.07%	0.26%	**0.45%**	**0.77%**	0.55%	1.12%	1.07%
Mid-cap Stks	**212%**	0.13%	0.14%	0.56%	**0.84%**	**1.73%**	1.30%	1.30%	1.26%
Small-cap Stks	**207%**	0.16%	0.28%	1.03%	**1.46%**	**2.85%**	2.33%	1.34%	1.31%
Value	**163%**	0.13%	0.12%	0.50%	**0.74%**	**1.21%**	0.77%	1.14%	1.11%
Blend	**144%**	0.12%	0.11%	0.41%	**0.66%**	**1.04%**	0.59%	1.07%	1.07%
Growth	**215%**	0.13%	0.15%	0.55%	**0.84%**	**1.84%**	1.32%	1.33%	1.28%
Large TNA	**147%**	0.11%	0.10%	0.65%	**0.86%**	**1.44%**	1.01%	0.94%	0.97%
Small TNA	**188%**	0.13%	0.14%	0.46%	**0.73%**	**1.44%**	0.87%	1.27%	1.23%

Table IV – Descriptive Statistics of the Fama-MacBeth Regressions Variables

The variables used in the Fama-Macbeth (1973) regressions (table V) are summarized here. The table reports the total number of observations, and the average total net assets (in millions of dollars) expense ratios, 4-factor alphas (computed using the factor coefficients estimated with monthly data over the previous 36 months), trading volume, commissions, spread, price impact, gross flows (inflows plus outflows reported in N-SAR filings) and the percent of total net assets held in cash. The average of these values is reported for every two years from 1995 to 2005.

	1995	1997	1999	2001	2003	2005
Observations per year	421	837	1503	2209	3743	3831
Total Net Assets ($mm)	$ 1,629	$ 1,470	$ 1,689	$ 1,266	$ 931	$ 1,480
Expense Ratio	1.16%	1.17%	1.19%	1.18%	1.28%	1.23%
4-Factor Alpha	-1.54%	-1.48%	-1.11%	-1.65%	-5.97%	-1.93%
Trading Volume	166%	143%	147%	146%	161%	139%
Commissions	0.13%	0.12%	0.11%	0.13%	0.14%	0.13%
Spread	0.27%	0.19%	0.13%	0.08%	0.05%	0.06%
Price Impact	0.93%	0.69%	0.56%	0.35%	0.18%	0.26%
Gross Flow	102%	117%	110%	84%	78%	69%
Cash	0.7%	2.7%	2.7%	2.7%	2.5%	2.2%

Table V – Fama-Macbeth Cross-Sectional Regression Results

Carhart (1997) 4-factor fund alphas are regressed on lagged values of the indicated regressors in a Fama-Macbeth (1973) framework. The regression is estimated quarterly from 1995 to 2005, generating 44 estimates for each coefficient. Regression 5 is restricted to those observations with flow data (fund matched N-SAR) and regressions 6 - 8 are restricted to those observations with commissions data (family matched N-SAR). Variables are annual, scaled by total net assets. SEC turnover is the minimum of purchases and sales. TradeCost is our measure of the annual trading costs of the fund. TRADESIZE is a dummy for large versus small relative trade size, where trade size is relative to the average of funds in the same market capitalization category and trade refers to portfolio holdings changes. "Flow Cost" is the annual trading cost attributed to flow. Three interactive dummy variables for soft dollar expenditures (SD_Sales, SD_Research, SD_Rebate) are used, drawn from the fund's N-SAR disclosure, reflecting disclosure about the factors used in selecting brokers (question 26). A value of 1 (0) indicates a yes (no) response. Standard errors account for autocorrelation via Newey-West (1987) with 3 lags.

	1	2	3	4	5	6	7	8
Intercept	0.058 (1.80)	0.060 (1.84)	0.058 (1.86)	0.033 (0.99)	0.035 (0.84)	0.042 (1.13)	0.036 (1.01)	0.040 (1.13)
Expense Ratio	-1.66 (-2.13)	-1.80 (-2.63)	-1.45 (-1.95)	-1.44 (-1.92)	-1.23 (-1.26)	-1.71 (-1.87)	-1.54 (-1.88)	-1.57 (-1.93)
SEC Turnover		0.0023 (0.60)						
TradeCost			-0.41 (-1.98)	0.38 (1.00)	0.74 (1.43)	0.13 (0.43)	0.42 (1.00)	0.21 (0.68)
TRADESIZE				0.00016 (0.04)	-0.0032 (-0.05)	-0.00023 (-0.06)	0.00062 (0.13)	0.0015 (0.38)
TRADESIZE*TradeCost				-0.88 (-2.73)	-1.05 (-1.97)	-0.67 (-2.65)	-0.72 (-2.65)	-0.74 (-2.92)
Log(Total Net Assets)	-0.0042 (-2.13)	-0.0043 (-2.21)	-0.0040 (-2.18)	-0.0023 (-1.10)	-0.0025 (-0.96)	-0.0025 (-1.12)	-0.0023 (-1.09)	-0.0026 (-1.23)
Flow*TradeCost					-0.75 (-2.16)			
SD_Sales						0.0020 (0.38)		
SD_Sales*TradeCost						-0.24 (-0.69)		
SD_Research							0.0011 (0.28)	
SD_Research*TradeCost							-0.24 (-0.88)	
SD_Rebate								0.014 (2.87)
SD_Rebate*TradeCost								-0.50 (-1.38)
Number of Cross-Sections	44	44	44	44	44	44	44	44
Average # of Observations	525	520	524	524	225	449	449	449
Average Adjusted R^2	2.22%	3.18%	3.45%	3.72%	4.73%	4.36%	3.72%	4.47%

Table VI - Trade Motive, Commissions and Trading Activity

Table reports coefficient estimates for two sets of regressions. Regressions 1 – 3 have dependent variable equal to the fund's brokerage commission (in basis points, 10000*brokerage commissions/(fund purchases+fund sales)). Regressions 4 – 6 have dependent variable equal to the fund's trading volume. The brokerage commission variable is calculated from data in the fund's N-SAR (question 21 and 23). Turnover is calculated using the semi-annual fund purchases, sales and portfolio total net assets variables from the N-SAR (question 71, parts A, B and C). Soft dollar disclosure is from N-SAR filing question number 26. Load indicator (N-SAR question 29) takes a value of 1 if the fund charges a load, 0 otherwise. Family Size is the log of the sum of TNA of all funds advised by the same management company (as identified in the CRSP mutual fund database). The net flow variable (%) is estimate as the change in TNA for the quarter, net of return effects. The regressions include fixed effects by year and investment objective fixed effects. Standard errors are clustered by fund. t-statistics are included in parentheses.

		Dependent Variable					
		Commission Rate			Trading Activity		
		1	2	3	4	5	6
Intercept		12.3	11.2	12	126	102.7	111
		(5.1)	(5.0)	(5.3)	(3.3)	(2.7)	(2.7)
Soft dollar disclosure:							
	Broker Sales of Fund Shares	0.66			15.1		
		(2.0)			(2.8)		
	Soft Dollars - Research		0.66			19.3	
			(1.4)			(3.1)	
	Soft Dollars - Rebated Commissions			1.05			17.3
				(3.1)			(2.3)
Expense Ratio		3.45	3.53	3.63	31.8	35.3	37.9
		(7.4)	(7.6)	(8.1)	(4.7)	(4.7)	(5.2)
Load Fee (Y=1)		-0.68	-0.59	-0.56	-16.6	-21.1	-20.7
		(1.9)	(1.6)	(1.6)	(-3.5)	(-3.6)	(-3.5)
log(Fund TNA)		0.31	0.31	0.31	-10.42	-10.1	-9.92
		(2.7)	(2.8)	(2.8)	(-2.1)	(-5.0)	(-4.8)
log(Family TNA)		-0.36	-0.34	-0.35	4.32	5.63	5.45
		(-3.6)	(-3.4)	(-3.6)	(3.6)	(3.6)	(3.5)
Net Flow					124.5	64.8	61.6
					(2.8)	(1.8)	(1.8)
R^2 (OLS)		15.9%	15.9%	16.1%	11.6%	11.9%	11.9%
Number of Obs.			18635			18239	
Fixed Effects	Investment Objective			Yes			
	Year			Yes			

Figure 1

The figures show annualized 4-factor alphas sorted by expense ratios (figure 1A), total net assets (figure 1B) , turnover or the minimum of fund purchases/sales divided by fund total net assets (figure 1C) and the estimate of annual trading costs (figure 1D). Each year, funds are sorted into quintiles based on the variable of interest. The average annualized 4-factor alpha (alpha is calculated using the 4 factor coefficients estimated over the previous 36 months) is then calculated for each quintile and reported in the figures. Quintile 1 represents the lowest value of each variable of interest and Quintile 5 represents the highest value.