The Journal of Wealth Management

Mutual Fund Economies of Scale: *Nature and Sources*

John A. Haslem

JWM 2017, 20 (1) 97-124
doi: https://doi.org/10.3905/jwm.2017.20.1.097
http://jwm.iijournals.com/content/20/1/97

Institutional Investor Journals
1120 Avenue of the Americas, 6th floor,
New York, NY 10036, Phone: +1 212-224-3589

Mutual Fund Economies of Scale: *Nature and Sources*

JOHN A. HASLEM

JOHN A. HASLEM
is professor emeritus of finance in the Robert H. Smith School of Business at the University of Maryland in College Park, MD.
jhaslem@rhsmith.umd.edu

"[P]revious studies show that 12b-1 plan expenses are deadweight costs and that these expenses do not reduce costs for shareholders."

—Malhotra, Martin, and Russel [2007]

In general terms, *firm* economies of scale provide cost reductions obtained from increasing asset size. Costs per unit of output decrease with increasing scale and constant variable costs as fixed costs are spread over more units of output. Operational efficiency is also often higher, with increasing scale leading to lower variable costs. In the case of *mutual funds*, there are fixed costs that go toward reducing the ratio of fund expenses to total net assets as assets increase.

Mutual funds are basically firms with inputs of financial and human capital. Fund outputs are sets of portfolio securities. It appears there are limits to human capital that can be productively added to funds. Research continues to explore these and other limits.

The Investment Company Institute (ICI), the fund industry trade association, provides a more inclusive statement of economies of scale:

> A *firm* having economies of scale is able to increase output with a less than proportional increase in labor and capital inputs by relying on efficiencies in the production process. Applying standard concepts of the firm to a financial organization, including a *mutual fund*, is not straightforward because of difficulty in defining and measuring output. Assets are the typical measure of output, but they may not capture the full range of services provided by a mutual fund. (Rea, Reid, and Millar [1999])

The ICI then describes economies of scale for *mutual funds*:

> Advisory and administrative fees are determined by contract and typically are computed as a percentage of assets.… The declining rate schedule reflects the expectation that cost efficiencies or scale economies will be realized in the management and administration of the fund's portfolio and operations as the fund grows. Such efficiencies do not generally arise from the spreading of fixed management costs across larger asset levels, as is often assumed. In fact, fund asset growth typically necessitates additional resources for portfolio management, investment research, and administration. Thus, the declining fee schedule results from anticipated efficiencies in the processes of the adviser and administrator as they add

> labor and capital to expand the scale of their operations. In this context, the contractual schedule reflects the expected long-run cost of managing and operating a fund at different asset levels, taking into account the necessary adjustments in scale made by the adviser and administrator. (Rea, Reid, and Millar [1999])

Finally, Gao and Livingston [2008] found that

> Economies of scale occur when the cost per unit of output decreases as output increases. These economies may result if a firm has fixed costs and constant variable costs. As the size of the firm increases, the fixed costs may be spread over more units and the cost per unit decreases. Economies may also occur if variable costs decrease as the size of output increases.

This study provides various findings on the nature and sources of mutual fund scale economies with respect to expenses, size, performance, trading, and numerous other effects. What remains is a general model of mutual fund economies of scale that identifies and incorporates all significant effects and their statistical significance and relationships.

COMPONENT EXPENSES AND ECONOMIES OF SCALE

In an early study, Latzko [1999] found that because many mutual fund expenses are fixed, asset growth should reduce the ratio of expense to average net assets. A translog cost function is used to assess evidence of fund economies of scale. The elasticity of fund total expenses relative to assets is significantly smaller than one, which indicates economies of scale. Although average costs decline over the range of fund sizes, at approximately $3.5 billion in assets the rapid decline in costs is exhausted.

Mutual fund total operating expenses include three major categories: management fees, 12b-1 fees, and "other" expenses. First, management fees are typically the largest fund fees paid to fund advisers, and it appears they do not grow proportionately to fund assets. These fees compensate fund advisers for providing fund services, such as security research and analysis. Average fund returns influence the size of management fees. High-performing fund managers are rewarded with higher compensation.

Second, 12b-1 fees that are paid for advertising, marketing, and distribution or sales fees are limited to 1.00% of assets, and they serve to attract fund assets. By attracting assets, 12b-1 fees enable economies of scale, but they also increase fund expense ratios. However, 12b-1 fees more than offset declines in front-end loads and increase fund operating expenses relative to growth in fund assets.

Third, account-based "other" expenses decline relative to mutual fund asset growth, and total expenses increase relative to asset growth. Other expenses include provision of statements and reports, dividends disbursement, custodial services, state and local taxes, auditing fees, legal fees, and director fees. These expenses have large economies of scale because of their fixed nature. The cost of servicing shareholder accounts is account-based and relatively fixed. Mutual fund costs are modeled as translog functions, which enable economies of scale to vary with asset size. Control variables include average expense ratios in the same investment objective, percentage annualized five-year returns, front-end loads, back-end loads, and fund assets in fund families.

Average mutual fund returns affect the size of management fees paid to fund managers. High-performing fund managers command higher fees than low performers. Brokers may be compensated by investor sales loads or from 12b-1 plan fees paid from fund assets. The relation among front-end loads, back-end loads, and expense ratios is complex because of load and no-load funds and share classes. Family funds tend to share expenses, thereby obtaining economies of scale unexplained by fund size. With larger family funds, potential economies of scale are larger and reductions in expense ratios greater.

In summary, because many mutual fund expenses are fixed costs, asset growth should reduce the size of the ratio of expenses to average net assets. Using a translog cost function and controlling for 22 investment objectives, the extent of economies of scale is estimated. The elasticity of fund costs is significantly smaller than one in all size categories, indicating economies of scale.

The average cost curve of the typical fund is downward sloping over the entire range of fund assets. Therefore, economies of scale are found in fund administration.

Latzko [1999] stated that

> the economies of scale in the mutual fund industry can be summarized by computing the

> average cost curve facing the typical mutual funds. This is derived by calculating the predicted average costs from the predicted total costs … for various fund asset sizes.… Average costs diminish over the full range of fund assets; however, the rapid decrease in average costs is exhausted by about $3.5 billion in fund assets.

ECONOMIES OF SCALE IN ADMINISTRATION

A later study by Latzko [2002] evaluated the existence and sources of scale economies in mutual fund administration. Past research is improved by using panel data, including up to seven annual observations on fund cross sections. Sources of fund economies of scale are found by examining the relationships between fund assets and various expense categories. The average elasticity of fund expenses relative to increasing fund total net assets is significantly smaller than one, indicating economies of scale. Average fund expenses are minimized at $22 billion in assets. The largest source of scale economies is "other" expenses, which are small portions of total costs.

Sources of mutual fund economies of scale are analyzed by examining relationships between fund assets and various expense categories. Management fees are the largest expense for most funds and exhibit small economies of scale. Management fees comprise a fixed percentages of assets, a declining rate in steps, or decline beyond stated asset breakpoints. Marketing and distribution fees also exhibit slight economies of scale in small funds. Other expenses, especially custodian, auditing, and legal fees, are reliable sources of scale economies. Analysis of fund management company profits indicates most or all economies of scale are passed on to fund shareholders in lower expense ratios, rather than being maintained as higher profits.

The percentages of mutual fund categories of operating expenses plus brokerage commissions include management: investment advisory fees (47.5%), brokerage commissions (12.8%), transfer agents (12.7%), marketing and distribution (11.0%), custodians (2.6%), auditing and legal services (2.1%), shareholder communication (1.6%), SEC registration (1.4%), and directors (0.6%).

Three major categories of mutual fund expenses are included in expense ratios. The first category is management fees that include investment advisory fees for security analysis and research services. Investment advisory fees compensate fund managers for expenses incurred in providing fund security research and analysis. Management fees do not seem to be significantly affected by asset growth.

Administrative costs may be included in management fees. Fund managers are rewarded with higher compensation for good performance. It may be more difficult and costly to manage large fund portfolios, but it appears unlikely that these costs grow at the same rate as fund assets, unless driven by costs of earning high returns.

The second category includes 12b-1 fees for mutual fund distribution and investor account servicing that are paid out of fund assets. When introduced, 12b-1 fees were primarily paid for advertising and marketing, but now their purpose is to increase fund assets by attracting more shareholders, which makes economies of scale possible. In fact, however, 12b-1 fees only add to fund expenses as *deadweight* costs.

The third category is "other" expenses for operating expenses and shareholder services. The major operating expenses are annual SEC registration fees. Custodial fees are paid to settle trades and hold fund assets as well as for accounting, auditing, and legal costs. Shareholder servicing costs are paid for providing shareholder communications, such as shareholder inquiries and printing and distribution of fund prospectuses and reports. Transfer agent fees are paid to maintain shareholder accounts, process transactions, and provide shareholder statements. Transfer agent fees are fixed per shareholder account, and average costs of maintaining shareholder accounts decline as fund assets increase.

Brokerage commissions are paid to brokers for mutual fund trades and are paid from assets. Funds also use commissions to pay for research services. These *soft dollars* are paid to brokers for providing fund managers with research services, including proprietary research, computer-related equipment, and software and databases. These costs are not included in fund expense ratios or reported in fund prospectuses or fund annual reports.

Global and international mutual funds are often more expensive to manage than are domestic equity and equity income funds, including expense ratios, custodian fees, management fees, and trading costs.

Mutual fund performance may affect the size of management fees paid to fund managers. Fund managers who generate higher performance may be paid higher compensation. Mutual fund sales representatives may be compensated from investor sales loads or fund

12b-1 fees. Funds with front-end or back-end loads may charge smaller 12b-1 fees. Funds with front-end loads have smaller expense ratios. The relations among front-end and back-end loads and expense ratios are complex because of fund issuance of multiple share classes with varying size expense ratios, front-end loads, and back-end loads.

Mutual funds in families tend to share expenses, such as those for computers, communication services, and shareholder accounting systems. As a result, family funds may have greater economies of scale than are explained by fund size. The larger the fund family assets, the greater the potential for fund scale economies and reduced expense ratios.

A translog function calculates elasticity of mutual fund costs with respect to assets. If cost elasticity is less (more) than one, fund expenses increase less (more) than is proportionate with assets, indicating economies (diseconomies) of scale. Overall, mutual funds have modest economies of scale. Fund average costs first rapidly decline, but then increase with fund size.

All average size categories of mutual funds indicate economies of scale. Cost elasticities vary across fund sizes and fall rapidly with asset growth. When brokerage commissions are included in total fund costs, very large funds have diseconomies of scale. All cost elasticities of fund investment objectives exhibit economies of scale and are similar for most objectives.

Management fees (including contractual investment advisory fees) compensate mutual fund managers for services provided. Investment services include identifying and analyzing investment opportunities consistent with fund objectives and policies, monitoring and reviewing portfolios, and determining amounts and timing of investment purchases and sales. There are small economies of scale for management fees.

If soft dollar arrangements reduce management fees, then controlling for trading volume, brokerage commissions should be negatively related to mutual fund investment advisory fees. Commissions are positive, however, but not significantly related to advisory fees. Thus, soft dollars do not reduce management fees.

Administrator fees are paid for oversight of performance of companies providing mutual fund services and may be paid out of management fees. High management fees are not related to lower other costs. Furthermore, high-cost funds have high costs in all expense categories.

Distribution fees finance activities that are intended to increase sales of mutual fund shares. Fees are spent on printing of prospectuses and reports for nonshareholders, preparation and distribution of advertising and sales literature, and payments to broker/dealers and shareholder servicing agents. Distribution fees are charged to current shareholders to fund efforts to attract new shareholders. Load funds have higher distribution fees than no-load funds.

Distribution fees are positively related to shareholder servicing expenses, which exhibit economies of scale. Distribution fees comprise some 30% of fund operating expenses. Distribution fees are 51% of total expenses for funds with back-end loads.

Distribution fees are higher for load mutual funds than no-load funds. Equity and foreign funds have higher distribution fees than bond and domestic funds. Distribution fees are positively related to asset ratios of both management fees and shareholder servicing costs. Distribution fees are not substitutes for other costs and have slight economies of scale, but they are deadweight costs to shareholders.

Shareholder servicing is the third largest mutual fund expense. Fund transfer agents maintain shareholder accounts, mail shareholder account statements and tax information, process shareholders' transactions, and calculate and dispense dividends. Some shareholder servicing agents respond to shareholder inquiries. Transfer agents are paid percentages of fund total net assets or by number of shareholder accounts. Load funds have higher shareholder servicing costs than no-load funds. Distribution fees are not substitutes for shareholder servicing fees, but they are positively related to these fees. Shareholder servicing expenses exhibit economies of scale.

"Other" operating expenses are the largest source of economies of scale. Cost categories are a small portion of total costs, but their cost elasticities are much smaller. Custodians physically maintain fund portfolio securities and safeguard cash and securities. Custodians are compensated based on percentages of net fund assets, but they may be paid by fees based on portfolio activity. Custodian fees do not increase with fund size. Equity funds have higher custodial fees than bond funds, and foreign funds have higher fees than domestic funds. Custodian fees have strong economies of scale.

Auditing and legal fees are the largest source of economies of scale and are highest for equity and foreign mutual funds. Fund family assets are negatively related

to fund-level auditing and legal fees. Funds provide shareholders with annual and semi-annual reports, involving printing and mailing expenses. Fund performance is negatively related to shareholder communication expenses, but loads are insignificantly related to these costs. Shareholder literature expenses, SEC registration fees, and director fees exhibit economies of scale. Director fees are higher for equity funds than bond funds, but smaller fees are related to higher fund performance. Fund family assets are negatively related to fund-level director fees.

In summary, the average mutual fund exhibits cost economies of scale. When brokerage commissions are included in fund expenses, average costs are minimized at $22 billion in total net assets, but larger funds have diseconomies of scale. All cost categories indicate economies of scale for average-sized funds, but distribution fees quickly exhibit diseconomies of scale as assets grow. Cost elasticities of auditing and legal fees become negative once fund assets increase sufficiently, but these fees decline once assets exceed $33 billion.

Actual economies of scale generated by mutual fund asset growth may be understated if fund managers do not pass all savings to shareholders. It is estimated that only 60% of total expenses paid by fund investors are spent on actual fund management. The remaining 40% of expenses goes to fund advisers. This issue is investigated using the operating profit margins of publicly traded mutual fund management companies. Fund revenues increase faster than costs, which may mean some economies of scale are appropriated by fund managers as operating income. This evidence is weak, however, because operating income elasticity is not larger than revenue elasticity. The evidence thus indicates that most, if not all, cost economies of scale result in lower expense ratios.

Latzko [2002] concluded that

> while some economies of scale are captured by fund advisers as higher profit margins, the conclusion is that most, if not all, of cost economies of scale from asset growth go to fund investors as lower expense ratios.

SIZE, PERFORMANCE, LIQUIDITY, AND ORGANIZATION

Chen et al. [2004] investigated the effect of scale on the performance of actively managed mutual funds for the years 1962–1999. Fund returns both before and after fees and expenses are found to decline with lagged fund size. This association is most pronounced among funds that invest in small and illiquid stocks and suggests that scale effects are related to liquidity. Controlling for size, fund returns do not decline with fund family size, which indicates that fund organization determines whether asset scale reduces performance. Data on single-manager and team-manager funds and composition of investments show that asset scale reduces performance as a result of the interaction of liquidity and organizational diseconomies. A negative relation is found between fund size and performance, especially for illiquid funds, but a positive relation is found between fund family size and fund performance.

Economies of scale are important in understanding the role of mutual funds. Scale economies provide insights into related issues that include importance of fund size to performance, fund inflows, persistence of fund performance, and fund manager–shareholder agency relationships. Data from 1962–1999 were used to determine whether fund performance depends on lagged fund size. Funds underperform market returns by 96 bps annually after fees and expenses, which is statistically significant. Fund average gross returns net of market returns are basically zero.

A fundamental issue in understanding mutual funds is economies of scale—how fund performance depends on fund asset size. This is an important issue for fund investors with the massive inflows that have increased fund assets. Persistence of fund performance also depends largely on the ability to scale fund holdings. Economies of scale also have implications for agency relationships between fund managers and investors and optimal manager compensation.

There may be advantages to scaling portfolio holdings, such as larger research resources and smaller expense ratios. However, large fund assets may decrease fund performance because of trading costs related to liquidity or trading price impacts. Small funds can easily invest in their best investment ideas. Lack of liquidity requires large funds to invest in less-than-best ideas with larger positions that decrease performance. Funds also have different styles (small-capitalization, value, momentum) and characteristics (portfolio turnover, age, expense ratios, loads, inflows, returns) that may drive performance.

There are other explanations that are consistent with negative relations between mutual fund size and

performance. The *liquidity hypothesis* argues that fund size reduces performance because of trading costs related to liquidity and price impact. If true, fund size should erode the returns of small-cap fund, which tend to be illiquid. In fact, fund size is much more important for other funds. Furthermore, size matters much more for the returns of small-cap funds than for other funds; in fact, size does not significantly affect other fund returns.

Liquidity, however, plays an important role in diseconomies of scale. Liquidity indicates that large mutual funds need to find more investment ideas than small funds do, but liquidity does not completely explain why large funds cannot scale stock holdings. Large funds likely can afford to hire more managers to cover more stocks. These funds can also take small positions in large numbers of stocks rather than large positions in a smaller number of stocks.

Size of mutual fund family is not clearly bad for performance. Controlling for fund size, assets of other family funds increase fund performance. The most likely reason for this is economies of scale in trading commissions and lending fees at the family level. Fund performance declines with size, but it increases with size of other family funds.

Mutual fund liquidity and scale need not be bad for fund performance, per se. Most family fund decisions are decentralized, including stock selection. Family funds may also economize on fixed assets. If large funds are organized as fund families with different managers, then scale need not be bad, and family size does not appear to be bad for family performance.

Mutual fund managers care a great deal about performance, and scale need not be bad for performance, per se. Large funds need more managers, and the organization of decision-making processes becomes important. Liquidity and scale likely erode performance because of certain organizational diseconomies. Many types of mutual fund organizational diseconomies lead to different predictions on why small funds outperform large funds. For funds with hierarchies, fund managers argue over implementation of ideas that affect choices of work efforts. Small funds are likely to outperform large funds at tasks processing *soft* (nondirectly verifiable) information. With soft information, fund managers have a harder time convincing others of their ideas. Soft information normally concerns research or ideas related to local stocks, including discussions with firm officials.

Large funds with hierarchies may spend too much on research efforts using quantitative measures to convince others or to implement ideas. Small-fund managers are much more likely to invest in local stocks, and they are also better at stock selection. Controlling for mutual fund size, funds with one portfolio manager are better at tasks involving soft information than funds with many managers. Single portfolio managers are significantly more likely than team managers to invest locally and to select better stocks than comanagers. Controlling for fund size, solo managers outperform team managers.

Hierarchy costs may not be present at the mutual fund family level as a result of commitments not to reallocate resources among family funds. Family funds have their own boards of directors, which means funds do not usually have to worry about fund families reallocating resources.

Several contributions are made here to mutual fund research. Fund performance declines with asset size, and the importance of liquidity in reducing this inverse relationship is established. Adverse effects of scale on performance are not inevitable because fund family size increases fund performance. Evidence is found that fund size and liquidity do erode performance and may be due to organizational diseconomies related to hierarchy costs.

Next, several dimensions of the findings are organized by topic. First, relationships between mutual fund size and performance show the following:

1. Gross fund returns are negative and statistically significant across four performance measures.
2. Fund family size gives a better prediction of performance than fund size.
3. Lagged fund returns suggest some persistence in performance.
4. Larger funds have lower expense ratios.
5. Fund loads and portfolio turnover proxy for active and passive funds.
6. Fund flow has little ability to predict fund returns.
7. Fund assets are negative and statistically significant across all measures of gross and net funds returns.
8. Fund age does not significantly predict fund returns.
9. Fund family assets are equally negative and statistically significant for both net fund returns and gross returns.

10. Fund size is negatively related to future performance and may be due to the following causes:
    a. transaction costs related to liquidity and price impact (*liquidity hypothesis*),
    b. large-fund investors being less discriminating about returns than small-fund investors because of greater advertising and marketing,
    c. small funds relying more on stock selection and performance than large funds to maintain investors (*clientele hypothesis*), or
    d. fund incentives locking in assets after periods of good past performance becoming passive (closet indexer) investors (*agency risk-taking hypothesis*).

Second, relationships among mutual fund liquidity, size, style, and number of portfolio stocks reveal the following:

1. Fund size matters more for performance in small-cap funds than large-cap funds because of liquidity issues.
2. Agency incentives vary little between small-cap funds and other funds.
3. The effect of fund size on performance is smaller for small-cap growth funds than for other funds.
4. Fund size does not affect gross returns and net returns of large-cap funds.
5. Fund liquidity erodes performance.
6. Large funds do not significantly scale up stock holdings or stocks covered relative to smaller funds.

Third, relationships between mutual fund family organization and performance reveal the following:

1. Controlling for family fund size, the size of other family funds increases fund performance—family asset size does not reduce performance, per se.
2. The effects of fund family size on fund performance are similar for small-cap growth funds and other funds.
3. Fund performance increases in large fund families.
4. Fund performance declines with asset size but increases with the size of other family funds, perhaps as a result of scale economies in family marketing.
5. Large fund families have huge economies of scale resulting from lower trade commissions and higher lending fees.
6. Family funds indicate that liquidity and scale are not necessarily bad for performance.
7. Large family funds are concerned with net returns and invest to maintain performance, which leads to more assets under management.

Fourth, relationships between mutual fund organizational diseconomies and performance reveal the following:

1. In addition to liquidity, fund organizational diseconomies (hierarchies) reduce performance.
2. Bureaucracy likely does not reduce fund performance with asset growth.
3. Small funds should outperform large funds with hierarchies in processing soft (non-directly verifiable) information.
4. Larger funds with hierarchies have issues in getting ideas implemented that decrease performance.
5. All else equal, small funds, hierarchies may outperform large funds.
6. The fund manager's, rather than the fund family's, control over funds reduces hierarchy costs.

Fifth, the relationships between mutual fund size and performance include the following:

1. Fund managers have the ability to invest locally and earn superior returns.
2. Small funds and small-cap funds are likely to invest higher proportions of portfolios in local stocks.
3. Family fund size does not influence investing in local stocks.
4. Performance differences in small funds and other funds are especially large in small-cap funds.
5. Large family funds do not perform better in local investing.

Sixth, relationships between mutual fund organization and investment composition reveal the following:

1. Controlling for fund size, funds managed by committees should invest less in local stocks.
2. Single-manager funds are significantly more likely to invest in local stocks than are team-managed funds.
3. Comanaged funds that invest locally perform the worst.

Seventh, relationships between mutual fund management structure and performance reveal the following:

1. Fund organization may be less predictable than fund size in predicting performance.
2. Effects of management structure on fund performance continue to hold.
3. Funds managed by committees underperform, but with smaller adverse effects than fund size.

In summary, there is a negative relationship between mutual fund alpha and past returns. On average, future fund alphas are smaller for large funds, and past returns are related to higher future alpha and performance predictability. Thus, there is strong evidence that fund size reduces performance. This scale relationship is not driven by heterogeneity in fund styles, size correlation with observable fund characteristics, or survivorship bias. The impact of fund size on returns is strongest for small-cap funds, in which liquidity issues significantly reduce performance. Organizational diseconomies related to fund hierarchy costs may also generate diseconomies of scale, but fund family size does not significantly reduce performance. Finally, data for single-managed and team-managed funds and portfolio composition show that organizational diseconomies affect asset size and performance relationships.

Chen et al. [2004] concluded the following:

> First, we carefully document that performance declines with fund size. Second, we establish the importance of liquidity in mediating this relationship. Third, we point that the adverse effect of scale on performance need not be inevitable because we find that family size actually improves fund performance. Finally, we provide some evidence that the reason fund size and liquidity does in fact erode performance may be due to organizational diseconomies related to hierarchy costs.

INDEX FUND BEHAVIOR AND ECONOMIES OF SCALE

Dowen and Mann [2004] analyzed *pure no-load* mutual funds and found that equity fund trading is negatively related to returns, but expense ratios are not significantly related to returns. Potential capital gains and tax cost ratios are positively related to returns. Funds exhibit both economies of scale and scope. Scope economies exist when fund adviser fixed costs are allocated over more than one fund. Individual investor returns are higher when investing in large funds in fund families.

There are several important findings for equity mutual funds:

1. Funds in large fund families have lower expense ratios.
2. A negative relation exists between portfolio turnover and performance.
3. More frequent trading is related to lower returns, even after controlling for assets.
4. There does not appear to be a relation between expense ratios and returns.
5. A positive and significant relation exists between mean capital gains exposure and returns.
6. A positive and significant relationship exists between tax costs and returns—returns produce tax consequences.
7. Fund families have negative and significant relationships between expense ratios and assets at the fund level and fund family level—possible economies of scale at fund manager and investor levels.
8. Economies of scope benefit fund managers and investors—costs decline as the variety of types of funds increases.

In summary, mutual fund behavior indicates that over time managers of larger funds and larger fund families generate higher returns at lower cost. Much of the performance difference is related to differences in fund portfolio objectives and perhaps the time period of the study. The fund industry is concentrated, with nearly 75% of equity assets and more than 65% of fixed-income assets held by funds in the largest size decile. Larger funds have the lowest tax cost ratios. For individual investors, it is clear that larger funds in large fund families are more likely to generate superior returns at lower cost.

Equity mutual fund managers who trade less tend to generate higher returns, but fixed-income managers who trade more also generate higher returns. This latter result may be due to greater predictability of returns and the use of duration and convexity models. Equity and fixed-income fund managers who produce better returns provide lower costs. Funds with aggressive growth have the highest costs. High-quality corporate bond funds and government-related adjustable-rate mortgage bonds have the lowest costs.

Mutual fund trading includes several costs that affect the profitability of trades. These costs include direct brokerage costs and bid–ask spreads, which affect trading efficiency. Trading is costly and should only be performed when advantageous to shareholders. Managers who generate lower returns may trade more in an effort to improve performance.

Equity mutual fund managers who generate higher returns have lower costs. The highest costs are those of aggressive growth stock funds. Funds with high past returns have significantly greater potential capital gains exposure due to appreciated securities. On average, potential capital gains exposure is negative in all fund classes and is often substantial. The SEC requires this disclosure to alert shareholders of potential future tax liabilities; however, investors may consider these liabilities as evidence of high past performance.

Mutual fund expense ratios are consistent with economies of scale at both fund and fund family levels. The cost structure of the fund industry is also consistent with economies of scope. Cost ratios for individual funds decline as the number of different types of funds controlled by asset managers increases.

Dowen and Mann [2004] found that

> controlling for fund returns and fund assets, there is a negative and significant relation between [the] expense ratio and … total assets under management by a particular fund family.… The result is consistent with the possibility that there are economies of scale at the asset manager level as well as the individual fund level.… Economies of scope are addressed … [and] the result is consistent with the idea that mutual fund managers and investors benefit from economies of scope. Costs decrease as the variety of types of funds increases.

EVIDENCE OF FUND ECONOMIES OF SCALE

A paper published in the *Strategic Insight Overview* by Strategic Insight Mutual Fund Research and Consulting [2004], a mutual fund research firm, reported that fund economies of scale are very much in evidence. In 2002, actively managed equity funds with over $1 billion in assets had weighted-average expense ratios of 0.95%, and those with assets up to $100 million had weighted-average expense ratios of 1.67%.

All mutual fund simple average ratios of advisory, administrative, and operational fees (excluding 12b-1 fees) were 0.97%, and weighted-average ratios were 0.60%. All domestic diversified equity fund simple average ratios of advisory, administrative, and operational fees (excluding 12b-1 fees) were 1.09%, and weighted-average ratios were 0.74%.

When all actively managed equity mutual funds (excluding index funds) are considered, economies of scale are also found. The weighted-average ratio of advisory/administrative fees (excluding 12b-1 fees) was 0.63%, the operational fee ratio was 0.23%, and the total cost ratio (excluding 12b-1 fees) was 0.84%. The weighted-average advisory/administrative fee ratio of the smallest funds is about 50% higher than for the largest funds, and their operational fee ratios are about twice as much as those for the largest funds.

Operating fee ratios decline significantly as mutual fund assets increase because of economies of scale from larger proportions of large shareholder accounts, larger initial shareholder purchases, and secondary shareholder investments. These events have translated to lower transfer agent fee ratios that are charged on a per-account basis.

Mutual fund contractual advisory fee breakpoints mandate lower fees at specified levels of fund assets. Thus, economies of scale in advisory fees are found across the fund industry. A fund's advisory fees are required to be identical across all its fund share classes.

Advisory fee ratios for mutual funds using unaffiliated subadvisors for portfolio management are generally higher than for non-subadvised funds. For all domestic large-cap funds, advisory/administrative fee ratios are 0.55% for nonadvised funds and 0.70% for subadvised funds.

In 2003, 78% of equity fund net inflows went to no-load funds, 22% to front-end load funds sold at net asset value, 10% to contingent deferred sales charges (CDSC) level-load funds, and 9% to outflows from CDSC back-end load funds. Mutual fund investor use of financial advisors has increased to approximately 80%. Most investors no longer pay point-of-sale commissions, but rather some form of annual advisor consulting fees. Many such investors purchase no-load funds or front-end load "A" class shares at net asset value, and financial advisors are paid by commissions other than point-of-sale commissions. Other investors purchase level-load fund classes with annual 12b-1 fees that are passed on to financial advisors. Fund managers suggest

less than 20% of sales are transacted with meaningful customer point-of-sale commission paid to financial advisors. Sales of *pay as you go* level-load fund classes sold through financial advisors have increased in popularity among regional and national brokers, but "B" class fund shares with CDSCs have significantly declined in broker sales because of criticisms and restrictions on their use. It is estimated that 75% of investors in funds served by financial advisors no longer pay sales commissions but instead compensate advisors through annual consulting fees.

There appear to be economies of scale in brokerage commissions to total assets for larger mutual funds. In 2002, for example, median annual brokerage commission to total asset were 10.5 bps for the 20 largest actively managed equity funds and ranged from a low of 2.5 bps to a high of 18.2 bps for the fund with the highest portfolio turnover. Higher turnover is clearly related to rising commission costs. Domestic index funds averaged less than one basis point.

Finally, for the 20 largest domestic, actively managed mutual funds, the weighted-average portfolio turnover ratio was 59%, and the median ratio was 27%. For many small funds with high turnover, brokerage commissions to total asset ratios were much higher than the 7–10 bps of the largest funds. Smaller funds and management companies may not be able to trade as cost-effectively as the very largest. Trading commissions to total assets at many funds may range 10–20 bps or more. It is estimated that trade bid–ask spreads to total assets for domestic large-cap funds are as high as 20 bps.

Strategic Insight Mutual Fund Research and Consulting [2004] found that

> the effective advisory/administrative fee ratios of the smallest funds are about 50% higher than those of the largest funds, and their operational ratios are about twice as much.... [O]perational fee ratios also fall dramatically as fund assets grow, partly because larger companies can deliver scale savings to their customers.

COST EFFICIENCIES: 12B-1 FEES, INSTITUTIONAL/RETAIL, AND INVESTMENT OBJECTIVES

Malhotra, Martin, and Russel [2007] evaluated determinants of mutual fund cost efficiencies for the years 1998–2003. Empirical results show that fund cost increases have been less than proportional to asset growth. Funds without 12b-1 plans are found to have greater scale economies than those with 12b-1 plans. Institutional funds are found to have larger scale economies than retail funds. Fund families that are more focused on investment objectives have lower fund management costs than those with diversified investment objectives.

Mutual fund economies of scale are important for three reasons. First, individual fund expenses have direct effects on investor returns. Assuming economies of scale, each increase in fund size decreases expenses and increases investor returns. Second, individual investors pay attention to fund expenses, and costs influence fund net flows. Third, with the rapid growth in fund assets and number of funds, mergers are fast becoming a new phenomenon. Mergers provide gains from lower costs of fund management due to economies of scale.

This study differs from previous research in four ways: First, consistency in economies of scale is analyzed over a longer period of time: six years. Second, the impact of fund focus on management costs and consequent economies of scale is analyzed. Third, previous research found that 12b-1 costs are deadweight costs for investors. Economies of scale are examined by classifying funds as either 12b-1 plan funds or non-12b-1 plan funds. Fourth, to study economies of scale, funds are categorized as retail or institutional.

The translog cost model is most commonly used for analyzing economies of scale. The model implicitly assumes a U-shaped average cost function. It is used here because it allows economies of scale to vary with the level of mutual fund assets. Fund output is defined as total assets under management, and total cost is total expenses. Total operating expenses are a function of fund total assets and control variables that affect levels of expenses.

The most common measure of operating efficiency in mutual fund economies of scale research is elasticity of costs relative to fund output. Economies of scale exist when the rate of output growth is larger than the rate of cost increase. Cost elasticity is measured by the percentage change in cost relative to the percentage change in fund assets. The existence and extent of economies of scale are measured by taking the first derivative of the translog cost function relative to fund assets. If elasticity is less than one, economies of scale are identified.

Empirical estimates of mutual fund cost functions reveal several findings: First, the natural logarithms of assets, fund age, deferred load, front-end load, brokerage availability, equity investment objective, turnover ratio, three-year annualized standard deviation, and 12b-1 plans have positive effects on fund costs. Fund focus and institutional status have negative effects on fund costs. Second, the natural logs of mutual fund assets have positive coefficient estimates that are all statistically significant, which implies positive cost elasticity as asset levels affect costs. Third, mutual fund age is positively related to costs: Holding size constant, older funds do not necessarily result in reduced costs of management. Fourth, family mutual funds do not significantly affect the costs of management, but concentrated funds have lower management costs. Reduced management costs are an important source of superior returns. Fifth, deferred loads and front-end loads have positive, statistically significant effects on mutual fund costs, and management costs are higher than for no-load funds. Sixth, the number of broker distribution channels has a positive significant coefficient. Controlling for mutual fund size, the larger the number of channels, the higher the management costs; investors may benefit from access to more funds, but they pay more in management costs. Seventh, institutional mutual funds have significantly lower (1–15 bps) management costs. Eighth, mutual fund three-year standard deviations have positive, significantly significant coefficients. High-cost fund managers tend to take more risks to increase performance to recover high management costs. Ninth, mutual funds with 12b-1 plans have positive and statistically significant coefficients. Funds with these plans have significantly higher management costs. Benefits of plan costs are not passed on to fund shareholders; funds with 12b-1 plans have much higher management costs than non-12b-1 plan funds. Tenth, in all cases, mutual fund average cost elasticity is positive and less than one. Cost elasticity significantly differs from one in each year. Thus, fund costs increase less than proportionately relative to increases in assets.

Mutual fund estimates of economies of scale for investment objectives reveal several findings: First, natural logarithms of assets, broker availability, deferred loads, 12b-1 plans, and three-year standard deviations have positive effects on fund management costs. Institutional funds have lower management costs across all 13 investment objectives. Second, there are statistically significant (less than one) mutual fund economies of scale for all investment objectives. Economies of scale are largest for funds with asset allocation investment objectives. Third, mutual funds both with and without 12b-1 plans have statistically significant economies of scale. Funds without 12b-1 plans have consistently larger economies of scale and higher returns. Fourth, economies of scale for institutional mutual funds are larger than for retail funds in four of six years, and they are equal in two years. On average, institutional funds have larger economies of scale.

Estimates of economies of scale by mutual fund size reveal several findings: First, natural logarithms of assets, fund age, equity funds, broker availability, deferred loads, portfolio turnover ratios, 12b-1 plan funds, and three-year standard deviations have positive effects on fund costs, but concentrated funds and institutional funds have negative impacts on fund costs in all size categories. Second, in all cases, mutual fund size subsets have statistically significant economies of scale. Fund assets over $450 million have the largest economies of scale.

In summary, mutual funds are tested for scale economies and the factors that influence management fees. From 1998 to 2003, fund cost elasticity is found to be less than one annually for the overall sample, thereby indicating industry economies of scale. Funds in large fund families do not have lower dollar costs of management; however, if families focus on just a few investment objectives, fund dollar costs are reduced. Furthermore, older funds do not have lower dollar costs of management. Funds distributed through larger numbers of channels have higher management costs. The same result exists for funds with high portfolio turnover that seek to time the market. Institutional funds and funds without 12b-1 plans, however, have lower management costs.

Mutual funds categorized by asset size, investment objective, use of 12b-1 plans, and institutional/retail funds all have annual cost elasticities less than one. On average, institutional funds have larger economies of scale than do retail funds. Funds without 12b-1 plans have larger economies of scale than funds with plans. Therefore, investors in funds with 12b-1 plans have failed to obtain the intended objectives of economies of scale from increased sales incentives.

Finally, mutual fund economies of scale are not constant from year to year. Therefore, decisions on fund mergers purported to achieve economies of scale should not be based on one year's data.

Malhotra, Martin, and Russel [2007] found that

> in all cases, there are significant economies of scale for the subsets according to size. For the six years in our sample, the largest economies of scale are for funds with an asset size of over $450 million.

TRADING COSTS AND DISECONOMIES OF SCALE

Edelen, Evans, and Kadlec [2007] noted that a study by Berk and Green [2004] argued that high-performing mutual fund inflows eliminate return persistence because of diminishing returns to scale. The current study examines trading costs as a source of fund diseconomies of scale. Annual trading costs are examined for a large sample of equity funds and found to be comparable in size to expense ratios. Trading costs have higher cross-sectional variation related to fund trade size and have increasingly negative effects on fund performance as relative trade sizes increase. Relative trade size also subsumes fund size in fund returns, which indicates trading costs are the primary source of diseconomies of scale. In a later paper, Edelen, Evans, and Kadlec [2013] discussed the effects of "invisible" trading costs on fund performance.

The hypothesis that mutual fund trading costs are the source of diseconomies of scale is tested over the period from 1995–2005. Annual trading costs are found to be 144 bps, and expense ratios, 123 bps. On average, funds do not recover trading costs—$1 of trading costs reduces fund assets by $0.41. This result hides variation in the impact of trading costs on fund performance attributable to trade size and motive.

Trading costs have an increasingly negative impact on mutual fund performance as *relative trade size* increases. Relative trade size (not the same as fund size) refers to average trade size relative for funds in the same market cap category. The relation for trading costs and fund returns is large (small) for funds with small (large) relative trade size. For example, $1 in trading costs increases fund assets by $0.40 for funds with small relative trade size and $0.80 for funds with large relative trade size. Thus, trading scale effects are a source of diminishing returns to scale for active fund management.

If organizational factors cause the negative relation between mutual fund performance and asset size, large funds should underperform small funds regardless of relative trade size. If diseconomies of scale in trading are the cause, large funds should underperform small funds only if they have large relative trade size.

For large relative trade sizes, $1 in trades reduces mutual fund assets by some $0.80. Fund managers appear to trade far beyond the point at which, with large relative trade sizes, value added exceeds transaction costs. There are two potential motives for excess trading: investor flow and soft dollars. Soft dollars include excess trading commissions paid to brokers to cover research provided to fund managers and to provide a mechanism to bundle other services with trading. Controlling for both flow-induced trading and soft-dollar trading is important in explaining the trade–performance relation, but neither fully explains excess trading. Other possible reasons for trading beyond cost recovery are the agency-cost signaling hypothesis or the possibility that large-fund managers are myopic to trading-cost handicaps.

Agency motives explain excess mutual fund trading that provides benefits to fund advisers. Agency costs occur when fund managers execute trades that reduce investor benefits from trades. Soft-dollar commissions are to be included in fund filings when additional broker sales of fund shares (distribution), fund receipts of research, or commission rebates to fund managers are considerations in selecting brokers.

The three soft-dollar motives are related to mutual fund trading volume, and all are significantly positive. Soft-dollar trades have higher commission rates, and they motivate fund managers to trade more. This incentive translates into trades and partially explains excess trading.

The three soft-dollar factors motivate mutual fund manager trades when investor costs are larger than trade benefits and fund advisers receive direct benefits. Each factor is related to elevated trading volume and small declines in fund returns. Fund managers are motivated to use these trades beyond the point of cost recovery. As with fund flow, controlling for soft-dollar motives provides evidence of excess trading by funds making relatively large trades.

Each soft-dollar motive correlates with higher commission rates but not with lower trade execution costs. Surprisingly, the research motive has the smallest impact on commission rates. The controversial commission reimbursement motive has the largest effect on commission rates. Overall, soft-dollar disclosure is related to higher per-unit trading costs.

Disclosure of soft-dollar motivation has a negative impact on trading costs. The negative relation is not significant, but it is more consistent with an agency cost interpretation than the alternative interpretation that soft dollars provide useful information. Controlling for soft dollars reveals that funds with large relative trade size trade beyond the point of cost recovery.

Trading costs replace fund size as determinates of returns. Trading is positively (negatively) related to fund size for funds with relatively small (large) average trade size. Scale effects in trading are the primary cause of diminishing returns to scale.

Average annual mutual fund trading volume is 181% of total net assets and varies by market capitalization and investment style categories. Per-unit trading costs include percentage brokerage commissions, effective spreads, and price impact. Average one-way trading cost is 76 bps, 146 bps for small-cap stocks, and 45 bps for large-cap stocks.

Consistent with diminishing returns to scale for trading costs, per-unit trading costs are about 40 bps higher for large relative mutual trade cost funds than for small relative trade cost funds. The difference is only 13 bps between large and small asset funds. Diseconomies of scale for trading are best measured by fund average trade size rather than fund asset size.

Average annual mutual fund trading costs are 144 bps, and annual expense ratios are 121 bps; the variation for trading costs is much larger than for expense ratios. Trading costs potentially explain more variation in fund returns than expense ratios.

Excess trading is motivated by investor flow and soft dollars. Controlling for flow-induced and soft-dollar trading is important in explaining the trading–performance relation, but neither provides a full explanation. Trading beyond the point of cost recovery may be caused by agency-signaling or by managers of large funds not fully understanding their handicap in trading costs. On average, trading costs have negative effects on fund performance. This impact is confined to three trade motives: scale diseconomies, operational trades (flow), and related trades such as soft-dollar trades. Mutual fund flow has a negative impact on returns and provides the most direct evidence that trading costs are the basis for this effect. Flow effects do not subsume conditional dependence of trading costs on relative trade size.

In summary, this article makes four empirical contributions. First, trading costs are larger than expense ratios and have a significant negative relation to mutual fund performance. Second, the negative effect of trading is largest for funds with relatively large average trade size, but trading does not negatively affect funds with relatively small trade size. Controlling for trade size, fund performance is not related to fund size. Trading costs are the largest source of diseconomies of scale. Third, flow-driven trades are significantly more costly than discretionary trades, which only partially explains the negative effects of trading on performance. Fourth, soft-dollar trades are related to much higher levels of trading and have a negative impact on fund performance.

Edelen, Evans, and Kadlec [2007] found that

> consistent with the hypothesis that diseconomies of scale 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.

COMPONENT EXPENSES AND ECONOMIES OF SCALE

Gao and Livingston [2008] reported studies indicating that mutual fund expense ratios decline as fund assets increase. To test this hypothesis, total fund expense ratios of actively managed domestic equity funds are decomposed into component fees. Most reductions in total expense ratios derive from minor fees paid to external service providers, and the large majority of reductions in expense ratios are derived from the smallest one-third of funds. Advisory fees are the largest component of expense ratios, and they are essentially constant for larger funds. Marketing expenses are the second largest component of expense ratios, which increase as fund assets grow. Previous studies have examined mutual fund total expense ratios and found they decline for larger funds. One interpretation states that economies of scale of larger funds are passed on as lower percentages of expense ratios.

Individual components of mutual fund expense ratios are examined. Reductions in components of larger funds are primarily found in minor expenses that include custodian fees, printing expenses, registration fees, director fees, auditing fees, legal fees, and other fees. Advisory fees are the largest component of expense ratios, but they decline minimally with asset growth. Marketing fees are the second largest component of expense ratios, and they increase with asset growth.

Most mutual fund economies of scale derive from minor expenses of the smallest one-third of funds. These expenses are provided by outside suppliers. Fund economies of scale from minor expenses are primarily exhausted when assets reach $60 million. Funds may charge smaller relative advisory fees as assets grow, but marketing expenses increase.

Findings from 1996–2004 dispel the idea that economies of scale are widespread in all mutual fund fees and in all fund sizes. The major source of economies of scale is the smallest one-third of funds for minor services usually purchased from external providers.

This study differs from previous research by examining all components of expense ratios. Component fees are compensation for fund service providers. Some services are more likely than others to provide economies of scale and decline with asset growth. Some expense ratio fees may significantly decline with asset growth. Furthermore, some service providers may be better situated to retain realized scale economies rather than pass them on to funds as reduced fees. Service provider retention of scale economies may be realized in less competitive markets or by better bargaining positions as fund-affiliated service providers.

Mutual fund advisory fees are paid to fund management companies for managing fund portfolios. Some funds have *step-down* fee structures that may provide economies of scale with asset growth. Many mutual fund expense ratio components are primarily fixed costs, such as auditing fees. Other components, however, are expected to have large variable costs, such as servicing agent fees. Servicing agent fees reflect direct costs of fund interaction with shareholders. As fund size increases, shareholder servicing costs increases, especially reflecting advances in technology and increased competition in providing services. Funds may be able to grow with more effective advertising, but these costs may not reflect economies of scale. However, marketing expenses may increase fund size and generate economies of scale in related expenses.

The mutual fund industry is monopolistically competitive. Funds compete on the basis of service (past performance) but not on fees. Fee levels may wary widely, but they are not primary factors in attracting shareholders. Some funds compete by providing low fees that appeal to informed investors, but the majority of funds compete for assets on the basis of performance. Because evidence shows funds do not consistently outperform, these funds appear to target less-sophisticated investors. Past fund performance has little ability to predict future performance, but fund percentage fees have to predict future net returns.

The major total expense ratio components (and estimated average percentages) of actively managed diversified domestic mutual funds from 1996–2004 are (1) advisory fees (65%), (2) servicing agent fees (12%–14%), (3) marketing fees (10%), (4) administrator fees (5%–13%), and (5) all other fees (5%).

Asset-weighted minor fees are much lower than equally weighted fees. Minor total expense ratio components are (1) custodian fees, (2) printing fees, (3) director fees, (4) SEC registration fees, (5) auditing fees, (6) legal fees, and (7) other fees.

Mutual fund total expense ratios are negative and significantly related to the logs of total net assets, indicating considerable economies of scale, 75% of which are from minor fees. A large proportion of minor fees are paid to nonaffiliated service providers. Nonaffiliated provider fees are driven by market competition. Many major fees (management fees) may not be influenced by market forces.

Mutual fund family size provides small economies of scale in expenses. Large family size helps individual funds to reduce fees; however, family size has little impact on major fees, which suggests economies of scale derive primarily from minor fees. This result may stem from family bargaining power with third-party service providers.

Mutual fund directors typically sit on several fund boards in single-fund families. Funds in large fund families appear to benefit from reduced expenses, but family size has little impact on major fees. This suggests family-level economies of scale primarily derive from minor fees. Payments to directors have a moderately positive relationship with fund expenses, suggesting independent directors receive higher total compensation and approve higher fund fees. Funds that are more difficult to manage may have both higher expenses and director compensation.

Analysis of mutual funds expenses shows that

1. Portfolio turnover rates are positively related to expenses.
2. Load funds have significantly higher expenses than no-load funds.
3. Increases over sample time are larger for major fees than for minor fees;

4. Fund assets are positively related to advisory fees and marketing fees but are negatively related to all other fees.
5. Larger funds have essentially the same percentage advisory fees and marketing fees as smaller funds.
6. Fund family size minimally affects major fee ratios (except for servicing agent fee ratios) but does affect several minor fee ratios.
7. Funds with higher loads have higher expenses, 50% of which derives from marketing fees.
8. Load funds have larger advisory fees and servicing agent fees.
9. Fund fees have positive time trends, except for registration fees, director fees, and legal fees.
10. Marketing fee ratios have significant diseconomies of scale in fund size.
11. Higher fund marketing fee ratios are positive and significantly related to net flows.
12. All minor fee ratios are negative and significantly related to fund size.
13. Total expenses and major fees are negatively and significantly related to fund sizes in all three terciles.
14. Advisory fees and marketing fees are negatively and significantly related to fund sizes in the first tercile.
15. Advisory fees and marketing fees are not negatively and significantly related to fund sizes in the second tercile.
16. Advisory fees are negatively and significantly related to fund sizes in the third tercile but not to marketing fees.

In summary, this study examines components of mutual funds' total expense ratios. More than 75% of economies of scale in fund expenses derive from minor expenses, many of which are provided by third-party suppliers. Advisory fees are the largest component and are essentially constant for larger funds, but the marketing fees component increases as fund assets grow. Observed economies of scale are driven primarily by the smallest one-third of funds. Finally, larger funds exhibit minimal economies of scale.

Gao and Livingston [2008] concluded that

> these findings dispel the notions that economies of scale are widespread for all of the components of mutual fund fees and for all fund sizes. Most studies have fitted one fee function for the expense ratio and for funds of all sizes. Our analysis shows that the major source of economies of scale is for the smallest one-third of funds for smaller services typically purchased from outside providers.

SCALING STRATEGIES AND DIMINISHING RETURNS

Pollet and Wilson [2008] stated that if actively managed equity mutual funds have diminishing returns to scale, they should change investment behaviors as total net assets grow. Although asset growth has little effect on average fund behavior, large funds and small-cap funds diversify their portfolios as assets increase. This increased diversification, especially for small-cap funds, is associated with higher performance. Fund family growth is related to the addition of new funds holding different stocks. Families with numerous funds diversify less rapidly as they grow, suggesting they may influence portfolio strategies.

The average mutual fund does not outperform the stock market, and only a few actively managed funds persistently outperform passive strategies, both of which suggest lack of manager skill. If this is so, why do actively managed funds manage so much money? Research by Berk and Green [2004] indicated that diminishing returns to scale reconcile lack of average fund outperformance and performance persistence with existence of fund manager skill. Flows to funds continue until the marginal dollar can no longer be invested profitably.

In the current study, the impact of asset growth on mutual fund investment behavior is investigated to identify constraints on asset growth. Regardless of whether fund performance is affected by diseconomies of scale in equilibrium, fund behavior should react to constraints imposed by growth.

If mutual funds obtain new inflows, should they research a larger set of investment ideas, hire new staff, and expand research capabilities, or continue—as feasible—to invest in given sets of stocks? First, funds overwhelmingly react to asset growth by increasing portfolio ownership of shares rather than by increasing numbers of present securities. Funds appear very reluctant to diversify with growth, but rather tend to acquire ever larger numbers of shares in stocks already owned. Ownership shares above 5% are common for large funds. Results appear to identify limits to scalability

(ownership costs) of fund portfolios, such as trade price impacts or liquidity constraints, as proximate causes of diminishing returns to scale.

Second, diversification is associated with higher monthly risk-adjusted returns. Small-cap mutual funds gain the most from diversification, controlling for fund and fund family size. Smaller funds outperform larger funds when investing in small-cap stocks. Results favor liquidity constraints as an explanation for why large-cap funds diversify more slowly in response to asset growth.

These results are consistent with two ways liquidity constraints can affect mutual fund returns. In the first case, fund managers have no ability to identify new investment opportunities once existing opportunities have been exhausted. All fund managers can do is go down their portfolio lists and select the next-best stocks without incurring prohibitive ownership costs. If some fund managers can add superior stocks with greater ease because they have better lists, liquidity constraints will not lower returns as much. If so, fund managers diversify optimally, which reveals management skill. Thus, diversification will be related to better performance, size adjusted, particularly with large liquidity constraints.

In the second case, some mutual fund managers are overconfident in their ability to select superior stocks, or they underestimate transaction costs. Again, diversification will be positively related to performance, especially if liquidity constraints are large. Overconfident fund managers, however, do not diversify optimally. In either case, mutual funds with high ownership costs will exhibit positive relations between diversification and subsequent performance, controlling for fund size. In contrast, funds less constrained by ownership costs, such as small funds, large-cap funds, or large family funds, will have weaker relationships.

In the third case, mutual fund family asset growth is associated with large increases in the number of family funds, especially when their individual funds are large. Fund family asset growth is related to large increases in number of family funds, especially when family funds are already large. Furthermore, family fund portfolios appear to differ from one another because the number of different stocks in family funds grows as rapidly as or more rapidly than the number of funds as family size increases. Therefore, fund family growth appears to be strongly related to additional investment ideas generated in new family funds rather than existing funds. This result is most pronounced in large fund families that are more likely to establish new funds.

In the fourth case, the number of newer mutual fund family funds has additional effects on individual fund response to asset growth. Whereas the average fund diversifies slowly in response to asset growth, funds with younger peers diversify even more slowly. At minimum, fund families do not appear to increase their capacity to generate additional assets in each family fund. In fact, fund families appear to influence investment behavior in the opposite direction by focusing on fewer stocks. Alternatively, fund families may assist in reducing liquidity constraints in individual funds by enabling combined family holdings of particular stocks to be traded at lower costs. These lower costs may also explain why large family funds diversify more slowly.

These results for mutual fund families are consistent with large family funds maintaining a large market share with a broad variety of funds. Each family fund portfolio is kept distinct from those of its younger peers even as they become very large. This behavior is evidence of fund family product proliferation. Because fund flows respond to marketing and advertising, fund families may prefer to establish new funds rather than hire new managers in existing funds for marketing purposes.

In summary, first, new results are presented on ways mutual fund portfolios are affected by growth in total net assets. Fund managers primarily increase share holdings of portfolio securities they own, which suggests that fund managers scale up existing holdings as assets increase. Fund managers are not interested in new investment ideas, except to compensate for liquidity constraints. Fund managers appear to remain focused on their few best bets. Second, the number of portfolio holdings increases at a slow rate in response to inflows. Mutual fund managers thus act as if they internalize slowly growing ownership costs. This diversification in response to growth is less pronounced for funds in large families and for larger-cap funds.

New evidence finds the proximate cause of mutual fund diminishing returns to scale is the inability to scale investment strategies as fund size increases. Funds diversify and scale less as they grow, and small-cap funds, large funds, and less-diversified funds respond more strongly. This behavior is consistent with limits to scalability being related to liquidity constraints. A fund behavior response to size growth is documented, rather

than just a linkage between fund size and other characteristics and returns.

A positive relationship between diversification and subsequent performance is found while controlling for mutual fund and family size. This relationship is stronger for small-cap funds, presumably because they are more constrained. Either fund managers diversify optimally and levels of diversification reveal their skills, or they diversify suboptimally.

Mutual fund family growth, especially for dominant fund families, is mainly associated with the addition of funds rather than expansion of the scope of activities of existing funds. Results indicate that increases in the number of funds are associated with increases in family total net assets, rather than by new funds. There is evidence that new family funds differ from existing family funds because of rapid diversification at the family level and slower portfolio diversification at fund levels. This finding of monotonically increasing fund family size indicates that marketing issues may be particularly important for large families. Product proliferation is widely understood as a strategy to preserve market share. Consistent with this possibility, funds in large families diversify less as they grow.

Alternatively, diseconomies of scale in large mutual fund hierarchy costs quickly reduce the marginal product of human capital to the point that fund managers do not add useful investment ideas.

Pollet and Wilson [2008] stated that

> the price impacts of large holdings are the necessary seed of diminishing returns to scale.... Price impact requires managers to deviate from perfect scaling by increasing the number of distinct holdings as fund TNA grows.

SIZE AND PERFORMANCE EROSION

Reuter and Zitzewitz [2010] stated that the mutual fund literature has provided two well-known findings. First, actively managed funds indicate little ability to persistently outperform their peers. Second, new fund money flows disproportionately into the actively managed funds with the highest past returns. Traditional interpretations of these findings state that fund managers are unskilled and fund investors unsophisticated. Relations between fund size and performance are likely endogenous, with size indirectly related to performance via other characteristics.

Berk and Green [2004] challenged these two findings by arguing that they are consistent with skilled mutual fund managers and diseconomies of scale in asset management. Rational investors chase performance to the point that expected future returns are equalized across funds. In equilibrium, more-skilled fund managers manage more assets, but because of diseconomies of scale they earn the same expected future returns as less-skilled managers. However, this view depends most importantly on the degree of diseconomies of scale in asset management.

The goal of the current study is to measure the causal effect of mutual fund size on performance. Separate identification of diseconomies of scale requires a test that causes an increase in fund size for reasons related to future returns only through diseconomies of scale. Small changes in fund returns can have discontinuous effects on fund flows with discontinuous effects on fund Morningstar ratings. Assuming fund manager skills vary continuously across Morningstar rating thresholds, the causal impacts of ratings on fund inflows can be identified. Because the source of these inflows is unrelated to fund manager skill, inflows can be used to identify the effects of fund size on performance.

The causal effects of mutual fund size on performance are identified by exploiting small differences in returns that can cause discrete changes in Morningstar ratings that, in turn, generate discrete changes in fund size. Regression discontinuity estimates find little evidence that fund performance declines with size.

Mutual funds that have differing numbers of Morningstar stars and almost identical numerical performance rankings are examined. Funds with more stars get greater investor inflows and have slightly better performance over the next six months, but performance is slightly worse over the next 12, 18, and 24 months; however, this does not diminish the predictability of performance.

The dominant issue is the effect of mutual fund size on performance predictability. Do funds get so large that performance declines with no positive alphas? Fund performance predictability is found to be statistically very significant. Average large fund alphas are 27 bps higher than those for smaller funds.

Mutual fund–level weighted-average Morningstar ratings are investigated. Funds with higher ratings tend to charge lower average fees, offer fewer share classes, and be less likely to charge loads. Consistent with investor

reactions to Morningstar ratings, funds with higher ratings receive higher net flows in the next 24 months. One-star funds underperform other funds over the next 24 months, but there is little difference in performance of other funds.

It is necessary to identify differences in mutual fund size that are unrelated to fund manager skills to measure the effects of size on performance. These differences would not appear if investors were rational and informed. Regression discontinuity exploits the fact that funds with past returns immediately above Morningstar rating thresholds obtain higher ratings than funds with ratings immediately below thresholds. To the extent investors place positive weights on Morningstar ratings, funds with risk-adjusted returns immediately above rating thresholds are likely to obtain significantly larger inflows than funds with risk-adjusted returns immediately below thresholds.

The various approaches taken in the data for tests for diseconomies include Morningstar ratings, asset classes, and where economies and diseconomies of scale are most likely. None of these approaches support the view that diseconomies of scale reduce fund performance.

In summary, Berk and Green [2004] argued that more-skilled mutual fund managers manage larger funds, but because of diseconomies of scale they earn the same expected returns as less-skilled managers. The prediction that fund managers will manage larger funds threatens evidence on diseconomies of scale.

In the current study, discrete changes in mutual fund flows that are related to discrete changes in Morningstar ratings identify flows that should only affect fund returns through diseconomies of scale. These findings are simply interpreted as meaning that fund managers are unskilled and investors are unsophisticated.

Berk and Green [2004] challenged these two interpretations. They argued that both interpretations are consistent with the combination of skilled managers with diseconomies of scale in asset management. Rational investors chase performance to the point that expected future returns are equalized across funds. In equilibrium, more-skilled managers manage more assets, but they earn the same expected future returns as less-skilled managers. Therefore, diseconomies of scale are associated with management of larger funds. The empirical relevance of these findings depends upon the degree of scale diseconomies in asset management. The current study's goal is to measure the effect of mutual fund size on performance. If fund size is endogenously related to expected future returns, size is uncorrelated with future returns in equilibrium and frustrates estimates of diseconomies of scale.

Discrete changes in mutual fund flows associated with discrete changes in Morningstar star ratings are used to identify flows that should only affect fund returns through diseconomies of scale. Small changes in fund returns can have discontinuous impacts on fund flows by way of affecting Morningstar star ratings. Funds just above the thresholds in Morningstar ratings receive incremental net flows over the next six months. Over the next 6–24 months, there is little evidence of diseconomies of scale. However, none of the estimates are significantly different from zero. Analysis of fund investment categories finds little evidence of diseconomies of scale. There is also little evidence to attribute estimates of fund performance persistence to diseconomies of scale.

Morningstar ratings tend to be larger for mutual funds and funds in larger families. Funds with higher ratings also tend to have lower average fees, fewer share classes, and no sales loads. Funds with higher Morningstar ratings receive higher net flows over the next 24 months. One-star funds underperform others over the next 24 months, but with little change in performance of other funds.

To determine the causal effects of mutual fund returns, variation in fund size that is uncorrelated to fund manager skill must be identified. If investors were perfectly rational, this variation could not be identified. Funds with past returns immediately above Morningstar ratings receive higher ratings than those with past returns immediately below thresholds. To the extent investors value Morningstar ratings, funds with risk-adjusted returns immediately above (below) thresholds likely receive significantly more (less) inflows.

Discrete changes in mutual fund flows related to discrete changes in Morningstar ratings are used to identify flows that should only affect fund returns through diseconomies of scale. Confidence intervals around estimates of scale diseconomies implied by exogenous fund flows are wider than implied by cross-sectional comparisons of large and small funds. Diseconomies of scale may be larger than previously estimated because very skilled fund managers are more likely to manage larger funds. However, scale diseconomies are not large enough to overturn general findings of unskilled fund

managers and unsophisticated investors. When successful fund managers obtain additional assets to manage, it may overwhelm their ability to perform at high levels—the *Peter Principle*.

Mutual funds operate in very competitive and efficient markets that should mitigate fund returns for all but the few highly skilled managers. Funds are constrained in their ability to "pay for performance," and they often lose managerial talent. Fund returns arguably have larger luck components, making it more complicated to infer fund manager skill. To the extent that funds are sold primarily to unsophisticated investors, reallocation of assets to more skilled fund managers is less efficient. Fixed costs in fund management may explain why there is little evidence of net diseconomies of scale.

Thus, endogeneity of mutual fund size and resulting difficulties in identifying its impact on performance are examined. To generate exogenous variations in fund size, discontinuity in fund flows across Morningstar star ratings is analyzed. Small variations in fund size can cause discrete changes in star ratings that produce changes in fund size. Use of discontinuity regression reveals no evidence of fund diseconomies of scale.

Reuter and Zitzewitz [2010] stated that

> our insight is that small changes in fund returns can have discontinuous impacts on fund flows through their impact on the fund's Morningstar rating.... Then, because this source of fund inflows is uncorrelated with manager skill (and other factors affecting future returns), we can use these inflows to identify the causal impact of fund size on fund performance. In other words, we are using small deviations from the rational behavior assumed in the Berk-Green model to measure the extent of diseconomies of scale.

ACTIVE MANAGEMENT, LEARNING, AND DECREASING RETURNS TO SCALE

Pastor and Stambaugh [2012] discussed findings that the popularity of mutual fund active management is not puzzling despite poor performance. The explanation features decreasing returns to scale. As industry size increases, fund manager ability to outperform passive benchmarks declines. At this size, investors who believe in decreasing returns to scale expect higher fund performance. These beliefs persist because large industry size causes investor learning about returns to scale to be slow. The active management fund industry should decrease only moderately if underperformance continues.

Active mutual fund management remains popular despite poor performance. Numerous studies have found active funds provide average returns significantly lower than passive benchmarks. The growth of index funds remains relatively modest. The popularity of active management is not puzzling despite its poor performance record because the active management industry has decreasing returns to scale. Fund manager ability to outperform passive benchmarks declines as industry size increases. As more money chases outperformance, stock prices are affected, and performance opportunities become more elusive. Decreasing returns to scale is equivalent to assuming mispricing is reduced as more money seeks to exploit it.

Decreasing returns to scale help us to understand the popularity of active mutual fund management. Investors are uncertain about industry's alpha but learn about it from realized returns. By observing negative performance, investors infer that the industry's return is lower than expected and decrease allocations to active management, which can be seen in the rapid growth of index funds. The decline in the relative size of actively managed funds to total funds has been modest, and it is consistent with cushioning from decreasing returns to scale. Investors infer expected returns are too low, and they know returns will increase if their proportion of actively managed funds is reduced. With decreasing returns to scale, past underperformance does not imply future underperformance, but it does imply that investors should allocate less to active management.

The equilibrium size of the mutual fund industry depends largely on the degree of competition among investors and fund managers. Without competition, equilibrium industry size maximizes expected profits. If investors compete and fund managers do not, all profits go to fund managers as alpha. In perfect competition, the fund industry earns zero expected profits with no abnormal fees and investors earn zero alpha.

Results indicate an inherent externality in actively managed mutual fund investing with decreasing returns to scale. When investors compete, they dilute returns by investing to the point that expected alpha is zero. Because of this externality, fund competition results in overabundance of active management relative to profit, thus maximizing size, which makes it easier to understand

why active management is so popular. If more active management implies less stock mispricing, then more competition also implies more efficient asset markets. These findings have important policy implications.

In a fully competitive market equilibrium, the size of actively managed funds is compared to total industry size. The ratio of actual active fund assets under management to total passive and active assets is 0.87. Rational investors in actively managed funds chose this allocation after updating previous beliefs about expected returns and returns to scale. Despite the history of returns of active management, the 87% investor allocation to actively managed funds is consistent with prior beliefs, but only as long as they feature decreasing returns to scale. The observed large size of actively managed funds can be rationalized by decreasing returns to scale.

In contrast, the popularity of actively managed mutual funds would seem quite puzzling assuming constant returns to scale. This assumption is routinely adopted, with constant alphas unrelated to industry size. The size of actively managed funds should be zero with constant returns to scale. With expected returns equal to zero, it would be undesirable for mean–variance investors to invest in active management.

The decline in investor response to underperformance of active mutual fund management is restrained by decreasing returns to scale. Investors know that by investing less in active management, future active returns will be larger. Because of decreasing returns to scale, the size of actively managed funds is likely to persist for many years.

Reconciliation of the "active management puzzle" of large, actively managed mutual funds with their poor performance history is a primary contribution. The other contribution is that investor learning is slow concerning decreasing returns to scale in active management. As investors update beliefs about expected fund returns, they adjust the relative sizes of actively managed funds. The extent to which investors learn is thus endogenous. Investors learn by how much they allocate to active management, but what they allocate affects how much they learn. Relative equilibrium size of active management varies over time, but fluctuations are significantly reduced by decreasing returns to scale. Variation in relative size of active management impedes learning about expected fund returns and scale economies.

Berk and Green [2004] assumed that individual mutual fund returns are decreasing in size. As investors update beliefs about each fund manager's skill, funds with positive performance records attract new money and grow in size. Funds with negative performance records experience withdrawals and reduce in size.

Actually, actively managed mutual funds have negative overall performance records, yet the active management industry remains large. This apparent "active-management puzzle" is approached by analyzing the aggregate size of the active management industry. Investors do not expect negative past fund performance to continue.

Model investors face endogeneity that limits learning about expected alphas and scale economies. As investors update their beliefs, they change the percentages of actively managed funds. The extent of investor learning is thus endogenous, and what they learn affects fund allocations, but what they allocate also affects how much they learn. The equilibrium percentages of actively managed funds vary with fluctuations significantly reduced by decreasing returns to scale. Reductions in fluctuations impede investor learning about expected alphas and scale economies even with long fund histories. The result is that initial investor beliefs about returns to scale persist for long periods.

In summary, a resolution is proposed to the puzzle of why mutual fund active management remains popular despite poor performance. Where investors and mutual fund managers compete, the large size of active management can be rationalized if investors believe funds have decreasing returns to scale. If investors believe returns to scale are constant, however, they will not invest in active management even if initially optimistic about portfolio manager skill.

With decreasing returns to scale, mutual fund investors adjust their allocations to achieve desired future returns, but following fund underperformance, proportional investor allocation to active management should be smaller, yet still sizable. Both predictions are consistent with the evidence that actively managed funds have underperformed passive benchmarks for four decades. Passive fund management has grown dramatically since the 1970s, but active management is still more popular. Active management is likely to remain large for many more years, even if its performance remains poor.

Mutual fund investors face endogenous factors that limit their learning—what they learn affects how much they allocate to active management, and what they allocate affects how much they learn. With endogeneity,

the equilibrium allocation tends to vary little over time and results in slow learning about the degree of returns to scale in active management. Initial investor beliefs about returns to scale thus affect their investment allocations for long periods of time. The inherent difficulty in estimating returns to scale calls for further empirical work. In addition to estimating returns to scale at the aggregate level, they could also be measured for various segments of active fund management.

Research could also explore additional aspects of mutual fund investor learning concerning parameters governing returns to scale. Those parameters are held constant for simplicity, but they could plausibly vary as a result of exogenous shocks such as market liquidity. In such a setting, parameter uncertainty would be refreshed periodically, further slowing the investor learning process. Continued research into decreasing returns to scale in active fund management will likely find non-decreasing returns.

Pastor and Stambaugh [2012] stated the following:

> We argue that the popularity of active management is not puzzling despite its poor track record. Key to this conclusion is to realize that the active management industry faces decreasing returns to scale: any fund manager's ability to outperform a passive benchmark declines as the industry's size increases. As more money chases opportunities to outperform, prices are impacted and such opportunities become more elusive.... In that case, our modeling of decreasing returns to scale is equivalent to assuming that mispricing is reduced as more money seeks to exploit it.

PREDICTING PERFORMANCE AND ECONOMIES OF SCALE

Elton, Gruber, and Blake [2012] found that past performance of large and larger mutual funds during 1999–2009 exhibit performance persistence for holding periods of up to three years. Funds that outperform index funds with equal risk can be identified. Expense ratios are smaller for large funds, and they decline as funds increase in size or perform well.

Berk and Green [2004] argued that past mutual fund performance should not predict future performance. Successful fund managers capture excess returns with larger expense ratios. Alternatively, funds will increase in size, and because of resulting diseconomies of scale, excess returns will disappear. Causal factors could include high transaction costs, organizational diseconomies, or purchases of poor-performing securities.

The primary focus of the current study is to examine the effects of mutual fund size on performance predictability. Fund performance is primarily measured to predict future performance. There are strong relations between fund performance and future cash flows, between risk-adjusted performance and future risk-adjusted returns, and ranking of funds by similarity of portfolios to those of successful portfolio managers.

Successful mutual fund portfolio managers may capture excess returns by charging higher expense ratios per dollar managed or excess returns, and performance predictability may decline as funds get larger because of diseconomies of scale from higher trading costs, organizational diseconomies, and adding underperforming securities. For this argument to hold, diseconomies of scale must be large enough to offset declines in expense ratios as funds increase in size.

Empirically, dollars of expenses and management fees are higher for larger mutual funds, but expense ratios and management fee percentages decline with fund size. Management fees indicate a small tendency to decline with fund size. Administrative fees have a large fixed component and are strongly negatively related to fund size. It is clear that expenses decline with fund size, and components other than management fees have the largest impacts.

The mutual fund expense ratio for each of three subsequent years declines for outperforming funds and increases for underperforming funds. With respect to changes in expense ratios due to changes in fund size or performance, expense ratios do not reduce predictability of performance. Because expense ratios decline with fund size, other costs that can increase with fund size must increase enough to increase overall costs.

Both large and small mutual funds indicate significant predictability of performance. Past alphas are significantly related to future alphas, but fund size is not. These results hold up when forecasting both two- and three-year holding periods. That alphas do not disappear for larger funds may be due to expense ratios declining with size, but also the ability to offset increase in trading costs and the need to hold larger numbers of securities with larger shares of fund family resources, more access to better traders and analysts, and more access to the fund family's best investment ideas.

In summary, mutual fund expense ratios and management fees decline with size and high performance. Management fees normally decline with size, but administrative costs are largely fixed. For predictability of performance to disappear, funds must grow with good returns and diseconomies of scale to erode performance. If so, large funds should not allow performance predictability. Both the largest and the smallest mutual funds exhibit significant performance predictability. Future fund alpha was significantly related to past alpha, but fund size was not significantly related to future alpha.

There are several possible explanations for why alpha may not disappear for large mutual funds. First, fund expense ratios decrease with asset size. Second, increases in transaction costs are offset and the need to invest a larger share of fund family resources. Third, large funds may have greater access to the best traders and analysts. Fourth, because fund family flows depend on the performance of the highest-performing funds, these funds may receive the best investment opportunities.

Elton, Gruber, and Blake [2012] found the following:

> A fund twice that performs well gets new cash flows and grows in size. Diseconomies of scale, whether caused by increased transaction costs, the acceptance of less profitable investments, organizational costs or other reasons mean that the skill embodied in past returns disappears and returns are not predictable.... If ... investors take time to reallocate funds or even to receive and process data, then growth in size takes place over time. Thus diseconomies of scale, to the extent they exist, will impact performance slowly over time. In this case predictability can exist although it should disappear over longer time periods. Furthermore, predictability should change as a function of fund size.... [T]hen we should find no predictability among big funds for which diseconomies of scale are more likely to be important.

INSTRUMENTAL VARIABLES AND SIZE AND PERFORMANCE RELATIONS

Phillips, Pukthuanthong, and Rau [2013] discussed previous research that has found mixed results concerning whether mutual fund size is negatively related to performance. One reason may be that fund size and performance are endogenous. A set of instrumental variables that influence fund size but are unrelated to performance is identified, but there is little evidence that size directly affects performance. An indirect relation is, however, manifested as a result of preferential allocation of investment strategies to smaller family funds.

Previous mutual fund research has been unable to establish conclusively whether fund size is negatively related to performance. This is important because the average fund manager appears unable to outperform passive fund benchmarks. Berk and Green [2004] argued that this is because funds with skilled managers attract higher flows than those managed by unskilled managers. If fund performance is negatively related to size, in equilibrium, both types of managers will generate similar expected future returns. The lack of fund outperformance among managers is thus not inconsistent with at least some managers being skilled. The crucial assumption is diseconomies of scale in fund performance.

In the current study, a set of instrumental variables is identified to control for any endogeneity bias by including fund characteristics as controls. Instrumental variables are based on *stale return–chasing behavior*, in which investors react strongly to lagged returns relating to the ends of reported holding periods. Changes in holding-period returns (HPRs) resulting from dropping end-returns from the samples are mechanical and only give perceptions of changed fund performance. These variables directly affect fund size but have no perceivable relation with fund performance.

Furthermore, these results form the intuition concerning instrumental variables. Investors note improvements in HPRs but are not aware the sources are stale negative end-returns dropping from the horizons of HPR calculations. These signals do not provide new information concerning fund performance or manager skills, but they do disproportionately increase allocations to funds from investors chasing stale performance. The results are exogenous increases in fund size unrelated to current performance. Funds with inflows typically increase portfolio holdings rather than diversifying. Thus, fund asset growth from inflows increases size and worsens diseconomies of scale to the extent they exist. Stale-performance chasing is a nearly ideal instrumental variable because it directly affects the endogenous variable fund size without any apparent relation to performance.

There is a possible nonlinear relation between mutual fund size and performance. This relation is reasonable if driven by fund trading costs because fund diseconomies of scale should increase monotonically. With linear progressions of stock prices and share depths in limit-order books, the price impacts of trading twice as many shares should be as double these larger trades "walk" further up the order books. With extremely large trades or highly illiquid assets, however, the relation between price impacts and trades sizes may be nonlinear. Increased fund inflows and smaller liquidity holdings increase trading costs and place price pressure on holdings that impedes fund performance. Furthermore, endogenous factors related to fund size may impact performance nonlinearly.

A test of nonlinearity by mutual fund size and performance reveals a negative relation in the largest size quintile. Fund family managers must decide how to allocate their best strategies across family funds. Some strategies are general, but most are specific to fund investment objectives with scale limitations. To minimize individual strategy trading price impacts, overall fund family strategies may have multiple substrategies for asset subsets.

In mutual fund families having multiple funds with the same investment objectives, new family investing ideas may go to smaller, more nimble funds (*fund favoritism*). If there are size differences across family funds, however, a better strategy may be a relatively larger proportion of overall strategies for smaller funds (*strategy rationing*). There is evidence of both mutual fund family strategies. Partitioning family funds by those with and without competitors in the same investment objectives, there is a significantly negative relation between fund size and performance for large funds with in-family competitors. Without in-family competitors, there is no evidence of relations between size and performance. Contrasting pairs of large and small family funds with common investment objectives, it is found that, on average, 73% of small-fund assets are also held by large funds, but only 34% of large-fund assets are held by small funds. This evidence suggests strategy rationing across funds. The unique holdings of large funds underperform those held by small funds.

Prior to SEC fair disclosure regulations, large mutual fund families had access to material, nonpublic information from investment banks. Marginal evidence during this period shows a positive relation between family size and gross returns. After regulation was imposed, little evidence can be found of relations between family size and net returns.

Partitioning mutual funds with and without in-family competitors during the SEC preregulatory disclosure period reveals that the positive relation between family size and performance is reversed for funds in the largest size quintile. This finding suggests that private information obtained by large fund families was channeled to smaller family funds. For funds with no within-family competitors, the relation between family size and performance is insignificant across all fund sizes. In the postregulatory period, the negative relation between family size and performance is confined to the largest funds with in-family competitors.

Overall, mutual fund size does not appear to affect performance directly through liquidity or trading costs. The relationship between fund size and performance appears to be nonlinear. The result is a significant negative relation between size and performance, but only in subsamples of large funds with smaller within-family competitors with the same investment objectives. Fund families appear to preferentially allocate their best investment strategies to smaller funds.

An analysis of mutual fund risk-adjusted gross and net returns and related findings is as follows:

1. The relationship between fund size and performance is significantly negative for gross and net returns.
2. There is strong evidence of investor performance chasing.
3. Funds in larger families exhibit marginally lower relative performance.
4. Significantly positive relations between family size and performance that existed in the pre-SEC disclosure regulation period reversed in the post-regulatory period.
5. Smaller fund families outperformed larger families in the postregulatory period.

An analysis of causal relations between mutual fund size and performance reveals the following:

1. There is no relation between investors' stale performance chasing or changes in Morningstar ratings and subsequent fund performance, except via fund size.
2. There is no significant relation between fund performance and investors' stale performance chasing for three measures of gross and net returns.

3. Stale performance chasing is usually higher for larger fund families and funds, and it is consistent with the higher advertising costs of larger families.
4. There is limited evidence that funds with higher amounts of stale returns chasing tend to have higher loads.
5. There is no evidence of higher Morningstar ratings and subsequent high fund performance.
6. There is little evidence of statistically significant differences in other fund characteristics across changes in Morningstar ratings, except that larger funds in larger fund families tend to obtain more upgrades and downgrades out of higher fund ratings.
7. Larger funds in larger families tend to charge lower fees.
8. Consistent with stale return chasing, funds with larger gross returns receive disproportionate flows in subsequent periods.
9. Larger funds tend to exhibit larger fund family and fund flows in prior periods.

An analysis of mutual fund size and performance reveals the following:

1. There is little evidence that size affects subsequent performance.
2. Negative relations between fund size and performance are due to endogenous relations between size and other characteristics that affect fund performance.
3. Other fund characteristics have limited predictive power for fund characteristics, except for persistence in performance (significantly positive relations between persistence and lagged gross returns) and significantly negative relations between lagged fund family size and fund size.
4. Small-cap funds with high liquidity loadings exhibit lower relative performance.
5. There is a typically significantly negative relation between fund portfolio liquidity and performance.
6. Larger funds with more illiquid stocks tend to underperform peers.
7. Relations between fund size and performance are insignificant, including illiquid funds due to linear relations between size and performance.

Mutual fund families preferentially allocate investment strategies across funds, and portfolio holdings are not unique between large and small within-family competitors. Small fund holdings are mirrored in large fund holdings; however, large funds also use additional strategies that, on average, significantly underperform those of small funds. Findings are consistent with fund families preferentially allocating better strategies to smaller funds, resulting in them comprising a relatively small proportion of strategies of larger family funds.

The results of this analysis are broadly consistent with preferential allocation of superior mutual fund family strategies to smaller funds that drive relations between size and performance. This evidence is consistent with shifting relations between fund family size and performance as a result of SEC fair disclosure regulations. However, the competitive advantages of larger fund families are smaller. Past preferential allocations of fund family private information to smaller family funds finds persistent negative relations between family size and fund performance across regulatory disclosure regimes for large funds with in-family competitors.

In summary, past research has found mixed evidence that mutual fund size is negatively related to performance. One reason may be that the relation between fund size and performance is endogenous, which would mean size is only indirectly related to performance via other fund characteristics. This study identifies a set of instrumental variables affect impact fund size, but not performance. These variables are based on stale return-chasing behavior, indicating investors react strongly to lagged returns. Here lagged returns relate to the ends of reported one-, three-, and five-year HPRs. Changes in HPRs are obtained when end-returns are mechanically dropped and provide perceptions of performance changes. These variables directly affect fund size but are not related to performance.

Use of instrument variables provides little evidence that fund size affects performance. There is also little evidence when illiquid funds are analyzed or in the period beginning with SEC fair disclosure regulations. Thus, overall, fund size does not appear to affect performance directly through liquidity or trading costs. The impact in prior research appears to be driven by endogenous relations between size and performance, which appear to be nonlinear. Negative relations between size and performance in large funds with smaller within-family competition in the same investment objective suggest fund families give preference to smaller funds when allocating "best" investment strategies. This behavior results

in negative relations between size and performance in large fund families.

Phillips, Pukthuanthong, and Rau [2013] stated

> in sum, we find results broadly consistent with preferential allocation of superior strategies to smaller funds with fund families driving the relation between size and performance. Our evidence is also consistent with a structural shift in the relation between family size and performance coincidental with establishment of fair disclosure regulation by the SEC. However, our analysis suggests that the magnitude of the competitive advantage enjoyed by large fund families is smaller than previously documented. Additionally, preferential treatment of investment strategies derived from non-public information to smaller funds results in a persistent negative relation between family size and fund performance across regulatory regimes for large funds with within-family competitors.

ACTIVE MANAGEMENT AND RETURNS TO SCALE

Pastor, Stambaugh, and Taylor [2013] empirically analyzed returns to scale in actively managed mutual funds for the years 1979–2011. They found strong evidence of decreasing returns to scale. As the active fund industry increases in size, funds' ability to outperform passive benchmarks declines. On the other hand, estimates avoiding econometric biases do not find decreasing fund-level returns. Also, newer funds exhibit more manager skill that coincides with industry growth but preclude more skill from increasing industry performance. Performance declines over the typical fund's life and can be explained by industry growth and industry-level decreasing returns to scale.

Performance of actively managed mutual funds has been a longtime research interest. The extent to which actively managed funds can outperform passive benchmarks depends on fund manager skills in identifying investment opportunities, but also on various fund constraints such as decreasing returns to scale. If fund scale affects performance, scale and skill are interactive—that is, more-skilled small funds can outperform less-skilled large funds. To understand fund manager skill, it is necessary to understand scale effects.

Two hypotheses have been raised about mutual fund returns to scale. First, Berk and Green [2004] identified decreasing fund-level returns to scale. As the size of active funds increases, funds' ability to outperform benchmarks declines. Second, there are decreasing returns to scale at the industry level as fund size increases and ability to outperform declines. Both hypotheses are driven by liquidity constraints. At the fund level, larger fund trades have a larger impact on asset prices and reduce performance. At the industry level, more money chases opportunities to outperform and prices change, making these opportunities more elusive. Consistent with liquidity constraints, evidence is increasing that fund trading exerts meaningful price pressures in equity markets.

Whether returns to scale are at the mutual fund level or the industry level, or both, is not clear a priori. If all funds were to follow the same strategy, their performance would likely depend more on combined fund size than on individual size. If funds follow unrelated strategies, the opposite would hold. The merits of these two hypotheses must be tested empirically. The fund-level hypothesis has been tested, but to mixed results. This study provides the first test of the industry-level hypothesis. The fund-level hypothesis is also tested again with improved data and bias-free econometric methods.

Interaction between skill and scale in active mutual fund management is empirically tested from 1979–2011. There are two biases common to estimates of fund-level returns to scale. At the fund level, bias-free estimates do not reveal decreasing returns to scale, but rather provide strong evidence of decreasing returns to scale at the industry-level. The negative relation between industry size and fund performance is stronger with higher fund turnover and volatility and for small-cap funds.

Estimates of mutual fund-level decreasing returns to scale are found to be small. There is thus no reliable evidence of decreasing returns to scale at the fund level; however, there is consistent statistical evidence of decreasing returns to scale at the fund industry level. The negative relation between industry size and fund performance is stronger for funds with higher portfolio turnover and higher volatility and for smaller funds. These results appear to be sensible because funds that trade aggressively and those that trade less-liquid assets are likely to have higher price impact costs in a crowded industry.

Decreasing returns to scale have major implications for assessment of mutual fund manager skills.

Manager skills have increased over time from −5 bps per month in 1979 to 13 bps per month in 2011. Improvements in skills are more rapid for higher-skilled funds. Fund managers have therefore become more skilled as time progresses.

Increases in mutual fund manager skills have not, however, increased fund average benchmark-adjusted gross returns. Reconciliation of increased skill without increased performance combines industry-level decreasing returns to scale with steady growth in fund size. Industry growth has made it more difficult for fund managers to outperform, despite increasing skill. The fund industry is now much larger and more competitive than in years past and requires more manager skill just to keep up.

Increasing average mutual fund manager skill is not explained by increases within funds because the measure of skill is constant over time. Average funds entering the industry are more skilled than existing funds and typically outperform them. Funds aged up to three years outperform those aged more than 10 years by a statistically significant 0.9% annually. Funds aged between three and six years outperform the oldest funds, suggesting younger funds capture portions of their higher skill with higher fees.

The negative relation between mutual fund age and performance holds across funds and within funds. Erosion in fund performance appears to be driven by industry growth. As funds age, the fund industry continues to grow, and the entry of skilled competitors reduces fund performance. Controlling for fund size, the negative relation between fund age and performance disappears.

Results are thus consistent with several findings. New funds tend to be more skilled than older ones, perhaps as a result of more education or better application of new technology. Because of superior skill, new funds tend to outperform benchmarks and older funds, but as new funds age, performance declines from continued growth in industry size and skilled competitors.

Mutual fund skill is measured by gross alphas on the first dollars invested in funds with no competitors. Funds' ability to identify profitable investments is measured before they are eroded by decreasing returns to scale. Alpha and Sharpe ratios do not separate the effects of scale. The effort is to measure fund expected benchmark-adjusted returns while accounting for adverse scale effects.

Previous research indicates that mutual fund-level decreasing returns to scale are not pervasive across mutual funds. Negative size to performance relations appear to apply only to the least-liquid funds and perhaps those with large inflows. In the current study, endogeneity of fund size includes fixed effects to account for heterogeneity in fund manager skill.

There is no consistent evidence of mutual fund decreasing returns to scale. Estimates of effects of fund size on performance are economically small. These findings are not affected by industry size, sector size, family size, fund age, or fund turnover.

There is a statistically significant negative relation between fund industry size and fund-level performance. This relation is both economically and statistically significant and is consistent with decreasing returns to scale.

Mutual fund industry size reduces fund-level performance, especially for funds with high volatility and portfolio turnover. This negative relation is also marginally stronger for small-cap funds. Furthermore, funds with high turnover do not perform better, whereas those that trade more do tend to perform better.

Mutual fund managers have become more skilled over time, but this change has not increased fund performance. This finding is consistent with gradual growth in industry size, which adversely affects fund performance because of decreasing returns to scale.

Mutual fund performance declines with fund age and is almost monotonic up to 12 years. These results are not due to incubation bias, nor are they likely due to fund risk. New funds initially tend to outperform their benchmarks, but as they age and the industry continues to grow, more skilled new funds depress older fund performance.

In summary, the interaction between mutual fund manager skill and active fund management is empirically analyzed. Two econometric biases plague estimates of fund-level returns to scale, but bias-free estimates do not identify decreasing returns to scale at the fund level. However, there is strong evidence of decreasing returns to scale at the industry level. The negative relation between industry size and fund performance is stronger for funds with high portfolio turnover and higher volatility, as well as for small-cap funds.

Results on returns to scale shape the assessment of mutual fund manager skill. Skill is measured by fund gross alphas before erosion from returns to scale. New fund managers tend to be more skilled than existing managers, but despite higher performance, average fund performance has not improved. These two facts can

be reconciled by industry-level decreasing returns to scale. Observed growth in industry size has impeded fund-level performance despite improvements in fund manager skill. Increased industry size also helps explain findings that fund performance usually declines over its lifetime. Consistent with more-skilled new fund managers, young funds usually outperform older peers. As funds get older, however, performance tends to decline because of continued industry growth and the arrival of skilled manager competition.

Several new issues are raised. First, the upward trends in mutual fund manager skills may be driven by improved education levels or knowledge of new technologies, but no direct evidence is provided. Second, more information is needed on the mechanism by which industry size reduces fund-level performance. Third, despite new methodology, evidence on fund-level returns to scale remains inconclusive. Fourth, the econometric methods can be extended in several ways, such as allowing for variability in fund manager skills over time.

Pastor, Stambaugh, and Taylor [2013] stated that

> we find that funds have become more skilled over time, yet this improvement in skill has failed to boost performance. This evidence is consistent with the observed gradual growth in industry size, which has had an adverse effect on fund performance due to decreasing returns to scale.

CONCLUSION

The literature has been unable to provide a definitive model of the sources and nature of mutual fund scale economies and diseconomies. This study provides findings on the sources and nature of fund economies and diseconomies with respect to expenses, size, performance, trading, and numerous other factors.

The mutual fund discussions include the following select findings:

1. There are economies of scale in fund administration.
2. The average fund exhibits cost economies of scale.
3. Fund size greatly reduces performance.
4. Fund expense ratios are consistent with economies of scale at both fund and fund family levels.
5. Fund economies of scale are very much in evidence.
6. Economies of scale are primarily driven by the smallest one-third of funds.
7. On average, institutional funds have larger economies of scale than retail funds.
8. Scale effects in trading, rather than other factors, are the primary cause of decreasing returns to scale.
9. The proximate cause of fund diminishing returns to scale is inability to scale investment strategies as fund size increases.
10. Whether the focus is on preferred regression discontinuity estimates or estimates based on changes in Morningstar ratings, there is little evidence fund size erodes performance.
11. As industry size increases, fund manager ability to outperform passive benchmarks declines.
12. Expenses decline with fund size, with expenses other than management fees having the largest impacts.
13. Fund families preferentially allocate the best investment strategies to smaller funds, which results in negative relations of fund size to performance in the largest fund families.
14. Actively managed funds provide strong evidence of decreasing returns to scale.

REFERENCES

Berk, J.B., and R.C. Green. "Mutual Fund Flows and Performance in Rational Markets." *Journal of Political Economy*, Vol. 112, No. 6 (2004), pp. 1269-1295.

Chen, J., H. Hong, M. Huang, and J.D. Kubik. "Does Fund Size Erode Mutual Fund Performance? The Role of Liquidity and Organization." *American Economic Review*, Vol. 94, No. 5 (2004), pp. 1276-1302.

Dowen, R.J., and T. Mann. "Mutual Fund Performance, Management Behavior, and Investor Costs." *Financial Services Review*, Vol. 13, No. 1 (2004), pp. 79-91.

Edelen, R.M., R. Evans, and G.B. Kadlec. "Scale Effects in Mutual Fund Performance: The Role of Trading Costs." Working Paper Series, SSRN, March 2007. http://ssrn.com/abstract=951367.

——. "Shedding Light on 'Invisible' Trading Costs: Trading Costs and Mutual Fund Performance." *Financial Analysts Journal*, Vol. 69, No. 1 (2013), pp. 33-44.

Elton, E.J., M.J. Gruber, and C.R. Blake. "Does Mutual Fund Size Matter? The Relationship Between Size and Performance." *Review of Asset Pricing Studies*, Vol. 2, No. 1 (2012), pp. 31-55.

Gao, X., and M. Livingston. "The Components of Mutual Fund Fees." *Financial Markets, Institutions and Instruments*, Vol. 17, No. 3 (2008), pp. 197-223.

Latzko, D.A. "Economies of Scale in Mutual Fund Administration." *Journal of Financial Research*, Vol. 22, No. 3 (1999), pp. 331-339.

——. "Mutual Fund Expenses: An Econometric Investigation." Working paper, Pennsylvania State University, York, 2002.

Malhotra, D.K., R. Martin, and P. Russel. "Determinants of Cost Efficiencies in the Mutual Fund Industry." *Review of Financial Economics*, Vol. 16, No. 4 (2007), pp. 323-334.

Pastor, L., and R.F. Stambaugh. "On the Size of the Active Management Industry." *Journal of Political Economy*, Vol. 120, No. 4 (2012), pp. 740-781.

Pastor, L., R.F. Stambaugh, and L.A. Taylor. "Scale and Skill in Active Management." Working paper, August 2013, https://ssrn.com/abstract=2318788.

Phillips, B., K. Pukthuanthong, and P.R. Rau. "Size Doesn't Matter: Diseconomies of Scale in the Mutual Fund Industry." Working Paper Series, SSRN, August 2013, http://ssrn.com/abstract=2313218.

Pollet, J.M., and M. Wilson. "How Does Size Affect Mutual Fund Behavior." *The Journal of Finance*, Vol. 63, No. 6 (2008), pp. 2941-2969.

Rea, J.D., B.K. Reid, and K.W. Millar. "Operating Expense Ratios, Assets, and Economies of Scale in Equity Mutual Funds." Investment Company Institute, *Perspective*, Vol. 5, No. 5 (1999), pp. 1-15.

Reuter, J., and E. Zitzewitz. "How Much Does Size Erode Mutual Fund Performance? A Regression Discontinuity Approach." NBER Working Paper, September 2010.

Strategic Insight Mutual Fund Research and Consulting. "Mutual Fund Fees: Facts, Trends, Economies of Scale, and Market Forces." New York: *Strategic Insight Overview*, No. 1 (2004), pp. 1-17.