Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Empirical Finance 13 (2006) 249–273
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/jempfin

Timing and selectivity of mutual fund managers: An empirical test of the behavioral decision-making theory

Larry J. Prather [a,*], Karen L. Middleton [b]

[a] College of Business and Technology, East Tennessee State University, Department of Economics and Finance, Box 70686, Johnson City, TN 37614, USA
[b] College of Business, Texas A&M University-Corpus Christi, Department of Management, Corpus Christi, TX 78412, USA

Accepted 19 October 2005
Available online 3 May 2006

Abstract

Classical decision-making theory suggests that decisions made by an individual or a team of decision makers should lead to the same performance outcome. Conversely, behavioral decision-making theory argues that decisions made by teams result in superior micro or macro forecasts and performance outcomes. Our tests using mutual funds support the classical decision-making theory. The empirical results are time invariant and robust with respect to the selected index or model specification.

JEL classification: D7; G0
Keywords: Team decision making; Mutual fund performance

1. Introduction

The belief that managers may possess differential security selection and market timing ability has spawned a plethora of performance attribution studies. These are important for several reasons. First, superior performance challenges the strong form of the Efficient Market Theory (EMT). Second, investors want to avail themselves of expert money-management skills. If factors related to exceptional performance can be unearthed, these can be employed as screening criteria in investment decisions. Additionally, if determinants of performance are associated with

* Corresponding author. Tel.: +1 423 439 5668; fax: +1 423 439 8583.
E-mail addresses: prather@etsu.edu (L.J. Prather), Karen.Middleton@tamucc.edu (K.L. Middleton).

0927-5398/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jempfin.2005.10.002

managerial structure, they could serve as a source of competitive advantage for fund sponsors. Finally, the evaluation of portfolio manager performance is vital for investment companies in establishing compensation schemes to attract and retain top-quality managers. However, extant evidence on market timing and selectivity is equivocal on these issues.

Mutual fund investment decisions are frequently made not by individual managers, but by a team of managers. Teams have traditionally been defined as two or more people who interact and influence each other while achieving common objectives for which they are held mutually accountable (e.g., McShane and Von Glinow, 2000). While both individual managers and teams of managers have access to a cadre of experts who continually scan the environment for new information, the key difference is in the formulation of the final investment decision. If an individual manages a fund, the manager has the sole decision making responsibility. However, if a team manages a fund, the team must reach a consensus before enacting a decision. *Morningstar* shows that teams managed only 23 mutual funds in January 1982, but by January 2002, teams managed 1912 funds, suggesting that the importance of management teams is growing. Despite this proliferation of funds managed by teams, little research exists that addresses the differences in performance outcomes when the mutual fund is managed by a team of decision makers rather than by an individual decision maker. Mellers et al. (1998) assert that the classical decision-making theory and behavioral decision-making theory are at odds concerning anticipated performance outcomes. The classical decision-making theory suggests that decisions made by an individual or a team of decision makers should lead to the same performance outcome. Conversely, the behavioral decision-making theory argues that teams form different ex ante expectations, resulting in superior micro or macro forecasts and performance outcomes.

Much of the previous research regarding the relations between management structure and performance outcomes has been completed in a laboratory setting (e.g., Hogarth and Reder, 1987; McNamara and Bromiley, 1997; Zeckhauser, 1987). Missing from the literature are studies of experienced and knowledgeable decision makers and analysts in their prototypical work setting. Our research attempts to fill that void by using field data on mutual funds managed by both individuals and teams.

We contribute to the literature by testing the selectivity and market timing skills of mutual funds managed by an individual manager against similar funds managed by a management team. The distinction between timing and selectivity permits dissecting elements of the decision making process to see if teams matter more for one or the other. Our investigation employs multiple market proxies and time-periods, and both conditional and unconditional timing models. We find no statistical evidence that funds managed by a team perform differently than funds managed by an individual manager, which supports the classical decision making theory.

In Section 2, we survey studies relating managerial characteristics and performance and classical and behavioral decision-making literatures. Section 3 discusses our data. Section 4 presents our methodology and empirical results and Section 5 concludes.

2. Literature review

2.1. Extant studies of managerial characteristics and mutual fund performance

Contemporary literature recognizes the merit of examining managerial characteristics in relation to performance. Hambrick and Mason (1984) argue that tenure, age and education have the potential to predict outcomes. Shukla and Singh (1994) find those equity fund managers holding the Chartered Financial Analyst (CFA®) designation outperform equity fund managers

that are not CFAs. Additionally, managers from schools requiring the highest composite SAT scores (e.g., Chevalier and Ellison, 1999a) and managers with MBA degrees achieve higher returns, suggesting the importance of education (e.g., Chevalier and Ellison, 1999a; Golec, 1996). Golec (1996) finds that experience positively correlates with performance, but Chevalier and Ellison (1999b) find that the career concerns of younger managers elicit greater effort from them since termination is more performance-sensitive for younger managers. Porter and Trifts (1998) suggest that experienced managers may become complacent, resulting in a negative impact on performance. Gottesman and Morey (in press) extend previous studies in this area and find that managers from schools with higher average GMAT scores and managers from MBA programs ranked in the top 30 by *Business Week* exhibit better performance than managers from other MBA programs. Moreover, the CFA designation or a PhD was unrelated to performance.

Apart from managerial characteristics, managerial structure may also influence the security selection and asset allocation decisions of a mutual fund. Classical decision-making theory and behavioral decision-making theory are at odds concerning the impact of managerial structure on performance outcomes (Mellers et al., 1998). Classical decision-making theory argues that timing and selectivity decisions made by either an individual fund manager with ultimate decision-making power or a team of managers achieving a consensus decision would be the same. Behavioral decision-making theory suggests that the collection of information and reliance on team members' expertise can engender procedural rationality where new perspectives can be fully examined and freely promoted, resulting in above-normal performance outcomes. The goal of this paper is to compare security selection and timing ability with respect to management structure.

2.2. The classical decision-making theory perspective

Classical decision-making theory is founded on a rational choice model that assumes the actions of individuals taken in a highly competitive environment are compelled by market forces to become effective or to withdraw (e.g., Hogarth and Reder, 1987). The rational choice model combines beliefs (probabilistic estimates of the likelihood of the occurrence of various states of nature of the economy and asset returns given the realization of each state) and preferences (utility) to reach an optimal decision (e.g., Hollenbeck et al., 1995; Kahneman and Tversky, 1984). Thus, classical decision-making research focuses on the rational choices the decision maker should make to optimize outcomes (e.g., Hogarth and Reder, 1987; Kameda and Davis, 1990). Given the same information, the conditional expectations of both an individual decision maker and those of a team of decision makers would be identical.

Recent studies question the validity of the classical decision-making theory, suggesting that investors exhibit cognitive errors such as myopic loss aversion (e.g., Thaler, 1999), overconfidence (e.g., Daniel et al., 1998; Daniel and Titman, 1999; Odean, 1998), and moderated confidence (e.g., Bloomfield et al., 2000). In addition, investors acting in response to biased expectations can influence market prices (e.g., Barber and Odean, 1999; Olsen, 1996; Pressman, 1998).

2.3. The behavioral decision-making theory perspective

Both permanent and temporary groups have become essential tools for coping with uncertainty (e.g., McNamara and Bromiley, 1997). A cognitively complex judgmental task such as achieving a group consensus on an investment decision often requires integrating others' perspectives.

Table 1
Growth of mutual funds

Year	Number of team-managed funds	Total number of funds	Percentage of funds managed by teams
1982	18	449	4.01%
1983	18	482	3.73%
1984	29	559	5.19%
1985	34	680	5.00%
1986	42	836	5.02%
1987	50	1036	4.83%
1988	71	1293	5.49%
1989	83	1449	5.73%
1990	109	1577	6.91%
1991	126	1771	7.11%
1992	174	2075	8.39%
1993	266	2656	10.02%
1994	370	3676	10.07%
1995	557	4961	11.23%
1996	649	5936	10.93%
1997	794	7161	11.09%
1998	984	8693	11.32%
1999	1185	10,079	11.76%
2000	1314	11,159	11.78%
2001	1411	12,287	11.48%
2002	1912	13,772	13.88%

Growth rate in the number of mutual funds between January 1982 and January 2002. Columns 1 through 4 list the year, the total number of team-managed funds in existence at year end, the total number of funds in existence at year end, and the percentage of funds managed by teams.

Accordingly, the behavioral decision-making model advocates the introduction of value-expressive characteristics, for example, frame susceptibility and varying attitudes toward risk, into the decision making process (e.g., Kahneman and Tversky, 1979).

Initial research in behavioral decision making implied that the superiority of groups over individuals was largely based on the pooling and integration of information to form a solution, while later research began to focus on the intellectual and cognitive tasks performed by the group (e.g., Hinsz et al., 1997). Other findings suggest the potential for learning from others may be enhanced in a group setting when members with differing backgrounds, cultures, or reference groups are integrated (e.g., Bikhchandani et al., 1998; Hinsz et al., 1997). Not only is a team more comprehensive in its information gathering, groups use information-processing rules and strategies more consistently than do individual decision makers. Groups appear to incorporate feedback into decisions more effectively than individuals (e.g., Hinsz et al., 1997). Groups recall and recognize relevant information with less variability than individuals (e.g., Sniezek and Henry, 1989), and are able to recall a larger quantity of information (e.g., Vollrath et al., 1989).

3. Data and data screening

3.1. Data

We utilize the *Morningstar Principia Pro*® database to cull the initial sample. Table 1 presents the number of funds in existence at the end of each year and the number of those funds that are team managed. Return data for funds in continuous operation during the period January 1992

Table 2
Sample composition

Investment objective	Number of team-managed	Number of ind. managed	Average total assets		Average age	
			Team	Individual	Team	Individual
Large value	20	58	529.12	1497.53*	208.9	299.1*
Large blend	14	57	1460.70	4127.87	183.2	302.2***
Midcap	9	25	630.89	558.15	207.9	320.9*
Total	43	140				

Column 1 lists the *Morningstar* investment objective. The number of team-managed funds and funds managed by an individual that were in continuous existence during our sample period from January 1, 1992 through December 31, 2001 are presented in columns 2 and 3, respectively. The average total assets (in millions) and average age (in months) of funds for each investment objective classification are presented in columns 4 and 5, respectively. Both measurements are at the end of our sample period. Levene tests reveal that the appropriate tests for the equality of means of age and size are the heteroscedastic *t*-tests.
*, **, *** denote significant differences at the 10%, 5% and 1% levels, respectively.

through December 2001 are from the *Wiesenberger*® InvestmentView database.[1] We utilized the CRSP Survivor Bias Free Mutual Fund database to identify funds that "died" during our sample period and to extract their returns. We will discuss the issue of survivorship bias in Section 3.2.

We consulted the *Morningstar* database to establish the management structure. Management structure is described in several ways. For funds managed by an individual, or where a single individual is the primary decision maker of a hierarchical team with distributed expertise, the manager is reported by name. We classify these as managed by an individual. Where two or more managers manage together and the team approach is strongly promoted *Morningstar* reports "management team" in the manager section. We classify these funds as team managed.[2] Our sample of team-managed funds reveals that teams vary in composition. For our large blend sample, the maximum number of managers is 14 with an average of 6.6.[3] Both large value and midcap fund samples have a maximum of 11 managers with averages of 4.6 and 4.8, respectively.

Next, we sort the funds by the *Morningstar* investment objectives. Restricting the analysis to groups of funds within selected investment objectives is important for several reasons. First, funds sharing the same objectives have similar characteristics (e.g., Klemkosky, 1976; Sharpe, 1966). Additionally, controlling for characteristics helps measure performance (e.g., Brown and Goetzmann, 1997; Kothari and Warner, 1997). Funds within an objective class may share more similar loadings on macro factors (e.g., Christopherson et al., 1998; Ferson and Schadt, 1996; Ferson and Warther, 1996), more similar loadings on asset class proxies (e.g., Bello and Janjigian, 1997), and a more similar composition of option-like securities (e.g., Jagannathan and Korajczyk, 1986).

[1] Our selection of 10 years is motivated by the desire to have ample power to detect performance differences if they exist. The Association for Investment Management and Research (AIMR) Performance Presentation Standards Handbook (1997, p. 69) states, "A 10-year performance record (or a record for the period since firm inception if inception is less than 10 years) must be presented." We discuss the ineluctable biases that arise later.

[2] *Morningstar* also lists the names of managers when more than one manager manages the fund but the fund does not qualify as team-managed. Multiple relationships are possible and some anecdotal evidence suggests that some fund companies may list multiple managers merely to de-link fund performance to a specific manager in an effort to mitigate the impact of manager turnover on subsequent flows into the fund. Therefore, to get the cleanest possible test, we delete the multiple manager funds.

[3] Another line of investigation would be differential performance among teams of different sizes. However, due to the vast differences in the number of managers that constitute a team, and the small sample size of funds managed by a team, we did not pursue this line of inquiry.

Table 3
Sample characteristics

Characteristic	LVI	LVT	LBI	LBT	MCI	MCT
Panel A: Macro asset allocation						
% Cash	4.2	6.4	5.5	2.0	7.0	4.0*
% U.S. stocks	89.3	88.7	88.8	95.4***	88.7	91.4
% Foreign stocks	4.0	2.3	3.2	1.8**	4.1	2.8
% Bonds	2.2	1.8	0.4	0.3	0.1	0.8
% Other	0.4	0.9	0.3	0.5	0.2	1.0
Panel B: Portfolio characteristics						
Number of holdings	76.0	77.3	150.6	272.0*	86.5	130.7
% of assets in top 10 holdings	36.5	35.2	36.3	28.4***	31.9	23.8**
Turnover	69.2	99.3	84.2	47.8**	149.2	158.6
Tax efficiency	77.5	73.7	79.3	82.3	63.9	75.9**
Potential capital gain exposure	11.4	7.2	1.8	5.9	−40.6	−19.4
P/E	13.8	24.7***	29.8	30.1	35.1	32.4*
P/B	4.0	4.1	5.5	5.7	5.7	5.3
P/Cash flow	13.3	13.4	18.0	18.1	23.7	22.4
P/Sales	2.2	2.2	3.4	3.6	5.8	5.2
12 month yield	0.7	1.2**	0.4	0.6	0.1	0.0
Panel C: Micro asset allocation						
% Utilities	5.4	7.0	1.7	2.8***	1.2	2.2
% Energy	9.7	10.5	6.2	7.0	3.4	3.2
% Financials	24.7	23.5	17.3	19.2	8.8	12.8*
% Industrial cyclical	15.9	14.3	10.0	11.7**	5.3	6.9
% Consumer durables	2.5	2.1	1.3	1.6	1.9	3.7**
% Consumer staples	4.9	6.8***	9.5	6.5*	2.3	2.8
% Services	13.0	12.9	14.3	11.7	18.0	13.6
% Retail	4.9	4.8	7.1	6.6	6.0	6.5
% Health	9.0	7.7	13.9	15.1	25.1	27.9
% Technology	10.1	9.3	17.2	18.0	27.9	20.5

Summary sample characteristics, as reported by *Morningstar*, at the end of our sample period (December 31, 2001) are provided. Column 1 is the selected portfolio characteristic and columns 2 through 7 present the breakdown by investment objective and management structure. *Morningstar* derives tax efficiency by dividing 1+after-tax (or tax-adjusted) returns by 1+pretax returns. *Morningstar* defines potential capital gain exposure as the amount of the fund's assets that would be subject to taxation if the fund were to liquidate today. The first two characters of the classification represent the investment objective (LV=large value, LB=large blend, MC=midcap) and the third character represents the management structure (T=team, I=individual). The sample sizes are LVI=58, LVT=20, LBI=57, LBT=14, MCI=25, MCT=9. We use the Levene test to ascertain whether the heteroskedastic or homoskedastic *t*-test is appropriate for comparing means for each of the characteristics.
*, **, *** denote significant differences at the 10%, 5% and 1% levels, respectively.

We eliminate index funds and funds that changed investment policies or management structures.[4] Table 2 presents the composition of our final sample.[5] A striking feature of our sample is that funds managed by an individual are typically larger, in terms of total assets, and

[4] The decision to purge funds that altered their management structures is driven by the desire to avoid a structural change. Because portfolio managers rely heavily on the internal research infrastructure of the fund company, it would be fascinating to compare performance before and after the change in management structure to isolate management's influence on performance. However, our small sample size precludes a meaningful analysis.

[5] We do not scrutinize other investment objectives because the number of funds meeting our criteria in those objective classifications is insufficient to perform meaningful statistical investigation.

older than funds managed by a team. Levene tests reveal that the appropriate tests for the equality of means of the age and size variables are heteroskedastic *t*-tests; and the *t*-tests reject the null hypotheses of equal age for each of the investment objectives. Only the large value funds managed by an individual are significantly larger than their team-managed counterparts.

The differential characteristics of the team-managed funds and their counterparts managed by an individual may reveal differences in the decision making process between the two management structures. The characteristics are presented in Table 3. Panel A provides the macro asset allocation for the groups of funds and indicates that fund holdings within each class are similar with the exception that team-managed midcap funds held more cash than their counterparts did, and large blend funds held more US stock and less foreign stock than their counterparts. Panel B presents the portfolio characteristics. The most striking differences between management structures are revealed by the holdings. Teams tend to hold more stocks than individuals and have a lower percentage of assets concentrated in the top 10 holdings. This might suggest that teams are bringing more information to bear on the portfolio decision and finding more good investments. Panel C presents a micro view of asset allocation, which suggests that while some differences exist, there are no discernable fundamental differences between the management structures.

Table 4 presents the average annual expenses. The summary statistics suggest that the average expenses of funds managed by an individual manager are higher than the expenses of funds managed by a team. Large value funds managed by an individual have total annual expenses of 33 basis points (bp) higher than their counterparts that are managed by a team. A two-tailed homoskedastic *t*-test rejects the hypothesis of equal expense ratios at the 0.05 level. Large blend funds managed by an individual have total annual expenses of 43 bp higher than their counterparts that are managed by a team and the difference is statistically significant at the 0.10 level. Midcap

Table 4
Sample funds average expense ratios

Investment objective	Expense category	Team	Individual	Levene *p*-value	*df*	Homoskedastic *t*-test *p*-value
Large value	Total	1.122 (0.412)	1.453 (0.647)	0.155	76	0.036
	12-b1	0.110 (0.151)	0.207 (0.293)			
	Exp.	1.012 (0.292)	1.246 (0.458)			
	Fld.	1.700 (2.665)	2.582 (2.739)			
	Bld.	0.250 (1.118)	0.379 (1.254)			
Large blend	Total	0.846 (0.476)	1.278 (0.801)	0.519	69	0.059
	12-b1	0.068 (0.137)	0.188 (0.264)			
	Exp.	0.779 (0.426)	1.091 (0.610)			
	Fld.	1.196 (2.380)	2.148 (2.530)			
	Bld.	0.000 (0.000)	0.245 (1.057)			
Midcap	Total	1.254 (0.387)	1.537 (0.617)	0.189	23	0.210
	12-b1	0.117 (0.179)	0.200 (0.276)			
	Exp.	1.137 (0.233)	1.337 (0.451)			
	Fld.	0.611 (1.833)	2.180 (2.742)			
	Bld.	0.000 (0.000)	0.600 (1.658)			

The *Morningstar* investment objectives are listed in column 1 and the expense categories are listed in column 2. Average 12-b1 fees (12-b1), annual expense ratios (Exp.), front loads (Fld.), back loads (Bld.), and total expense ratios (Total), for team-managed funds and funds managed by an individual are presented in columns 3 and 4, respectively. Total expense ratios exclude load fees. All expenses are from *Morningstar* at the end of our sample period, December 31, 2001, and are presented in percent. Standard deviations for each expense category are in parentheses. The sample sizes are LVI=58, LVT=20, LBI=57, LBT=14, MCI=25, MCT=9. The Levene *p*-value, degrees of freedom, and *p*-values from the homoskedastic *t*-test of the hypothesis that total expense ratios are equal are presented in columns 5 through 7.

funds managed by an individual have total annual expenses of 28 bp higher than their counterparts that are managed by a team; however, the differences in expenses are not statistically different at conventional levels. An examination of the components of total expenses reveals that 12-b1 fees only make up between 8 and 12 bp of the total expense differences. Operating expense differences explain most of the total expense differences.[6] Because team-managed funds have lower expenses, their monthly net returns should be higher than their counterparts ceteris paribus. This creates a small bias in favor of the behavioral decision-making theory.

3.2. Potential sources of bias

Despite our assiduous screening process, two potential types of bias may exist: survivorship bias and incubation bias. Brown et al. (1992) show that survivorship bias exists when extinct mutual funds are excluded when studying performance and Brown and Goetzmann's (1995) probit analysis intimates that poor performance increases the likelihood of disappearance.

To examine the impact of extinction on our sample of funds we searched the CRSP Survivorship-Bias Free Mutual Fund database and found that 125 funds with identifiable investment objectives, end dates, and managers died during our sample period. Of these, only 32 funds have Wiesenberger investment objectives of large growth, aggressive growth, growth and income, income, or total return. While these investment objectives do not mirror the *Morningstar* objectives that we utilize, they are the most likely candidates to be classified by *Morningstar* as large value, large blend, or midcap. However, only five of these funds commenced operation before the beginning of our sample period and thus would have been eliminated in our screening process. Of those five funds, an individual managed four of the funds, and one fund did not have a manager listed. In an attempt to properly classify these funds, we contacted *Morningstar*. However, the classification of these funds could not be determined.[7] Our conjecture is that neither of the total return funds (TR) would have fallen into any of our classifications. Of the remaining two funds, we believe that the large growth fund would be classified into one of our large categories (LV or LB) and that the aggressive growth (AG) fund would be classified as a midcap or small cap fund. Moreover, the time series of these dead funds is relatively short with 44, 20, 50, and 63 monthly returns available during our 120-month sample period, respectively. Because efforts to properly classify these funds failed, including their returns in our analysis may induce some degree of bias.[8] Given the large relative size of our sample compared with the small number of dead funds, excluding these funds should not materially influence our results.

Arteaga et al. (1998) find that return bias can be created through mutual fund incubation. They argue that a fund sponsor could create multiple funds, each taking multiple successive bets in an uncertain world. After the outcomes are known, winning funds would be marketed and losing funds dropped which would function to augment the returns of managed portfolios relative to the index. Evans (2005) also examines this issue and reports evidence of incubation. In one case, Evans reports an incubation period of 41 months. Edelen (1999) argues that fund sponsors can use

[6] Operating expenses consist of portfolio manager compensation and operating costs. However, our data does not permit scrutinizing the makeup of these costs. Thus, we cannot tell whether teams or individual managers command larger total compensation.

[7] A list of these funds is available from the corresponding author.

[8] We ran the market model and Treynor-Mazuy timing model to examine the performance of these defunct funds. Three of the four funds had an R^2 of less than 0.30 with both models and none of the selectivity or the timing coefficients of the funds that exhibited a poor fit were statistically significant. The remaining fund had an R^2 of 0.64 with both models and selectivity coefficients were negative and statistically significant at the 5% level.

seed money to preclude a fund from incurring normal transactions costs during its infancy and reports that this can bolster returns by 1.4% per year. Evans finds that post-incubation returns are high for the first year but drop to normal levels afterwards. Thus, our decision not to admit new funds into our sample prevents a bias in favor of team-managed funds that may arise because team-managed funds have grown more rapidly than funds managed by an individual.

3.3. Computation of returns

We compute monthly net holding period returns for each fund as,

$$r_{i,t} = \frac{\text{value}_{i,t} - \text{value}_{i,t-1}}{\text{value}_{i,t-1}} \tag{1}$$

where $r_{i,t}$ is the return on fund i during period t, and $\text{value}_{i,t}$ is the value of a hypothetical investment in fund i at time t. Values of hypothetical investments are net of annual expenses and 12-b1 fees and assume reinvestment of all distributions such as capital gains, dividends, and interest. We calculate returns for the individually managed and team-managed strata by summing the returns of the individual funds i, within the individually managed or team-managed management group m, for each investment objective group o, and computing their average monthly returns for each period t. This approach is analogous to the approach that Becker et al. (1999), Connor and Korajczyk (1991), and Ferson and Schadt (1996) use in their examinations of asset classes. Formally,

$$\frac{r_{m,o,t} = \sum_{i=1}^{n} r_{i,m,o,t}}{n_{m,o}}. \tag{2}$$

Our procedure results in six sets of equally weighted index returns (one for funds in each investment objective managed by an individual manager and one for funds in each investment objective managed by a team). We utilize equally weighted indexes to palliate the undue impact that a large fund could exert on the results if value-weighted indexes were used. Connor and Korajczyk (1991) point out that comparing indexes also results in a lower residual variance and more precise parameter estimates.

4. Methodology and empirical results

4.1. Performance evaluation

Jensen (1968) developed an early test of performance using ordinary least squares regression (OLS):

$$R_{\mathrm{p},t} = \alpha_{\mathrm{p}} + \beta_{\mathrm{p}}(R_{\mathrm{m},t}) + \varepsilon_{\mathrm{p},t} \tag{3}$$

where $R_{\mathrm{p},t}$ is the excess return on the portfolio, α_{p} is the excess risk-adjusted return, β_{p} is the systematic risk of the portfolio, $R_{\mathrm{m},t}$ is the market risk premium and $\varepsilon_{\mathrm{p},t}$ is a well-behaved error term. Under the joint hypotheses of the CAPM and the efficient market hypothesis (EMH), α should be zero.

However, performance evaluation has become a controversial issue and our quest to investigate both security selection and timing abilities further complicates the issue. Therefore,

we commence with the Jensen model as a first test of overall performance and then use the unconditional Henriksson and Merton (1981) model (HM) as a first test of market timing. We examine the robustness of the timing and selectivity results by utilizing the unconditional Treynor and Mazuy (1966) model (TM), unconditional Fama and French (1993) three-factor model (FF-3), conditional HM and TM models, conditional FF-3 model, and the conditional Carhart (1997) four-factor model.

We utilize the S&P 500 total return index, Center for Research in Security Prices (CRSP) equally weighted NYSE and Amex index (CRSP EW), and CRSP value-weighted (CRSP VW) NYSE and Amex index as alternative market proxies. Lehmann and Modest (1987) furnish empirical evidence that risk-adjusted performance is affected by the selection of the market proxy. Our risk-free proxy for CAPM-based tests is the 90-day T-bill return.

4.1.1. Unconditional models of timing and selectivity

Treynor and Mazuy (1966) illustrate that performance can be dichotomized into security selection and market timing components. If fund managers maintain constant risk, the characteristic line will be linear with scatter inversely related to diversification. In this case, α in Eq. (3) above may be a sufficient performance metric. Alternately, market-timing attempts cause curvature in the characteristic line and the curvature and scatter will depend on the proportion of correct guesses and the magnitude of the gambles. Treynor and Mazuy (TM) use a quadratic term to enhance the coefficient of determination of the regression and to separate market timing from security selection measures. The TM model is

$$R_{p,t} = \alpha_p + \beta_p(R_{m,t}) + \gamma_p(R_{m,t})^2 + \varepsilon_{p,t} \tag{4}$$

where $R_{p,t}$ is the excess return on the portfolio during period t, α_p measures selectivity (risk-adjusted return), β_p is the systematic risk of the portfolio, $R_{m,t}$ is the excess return on the market during period t, γ_p measures timing ability, $R^2_{m,t}$ is the squared excess return on the market during period t.

The Henriksson and Merton (1981) model replaces the squared risk premium with a variable max (0, R_m):

$$R_{p,t} = \alpha_p + \beta_p(R_{m,t}) + \gamma_p(R_{m,t})D_t + \varepsilon_{p,t} \tag{5}$$

where $D_t = 1$ if the market return exceeds the risk-free rate at time t.

4.1.2. Conditional models of timing and selectivity

Becker et al. (1999), Ferson and Schadt (1996), and Ferson and Warther (1996) suggest that a misspecification of unconditional timing models may explain poor selectivity and perverse timing since common time variation in risk and risk premiums are ignored. Assuming semi-strong form market efficiency, the authors incorporate vectors of lagged instrumental variables intended to capture public information. The public information vectors found individually and collectively significant vary across studies, but are related to the lagged level of the Treasury bill yield, lagged index dividend yield, and the lagged term structure slope.

Ferson and Schadt (1996) derive the conditional Treynor-Mazuy model as:

$$R_{p,t} = \alpha_p + b_p R_{m,t} + C_p'(Z_t R_{m,t}) + \gamma_{tmc}[R_{m,t}]^2 + \varepsilon_{p,t} \tag{6}$$

where $R_{p,t}$ is the excess return on the portfolio in period t, $R_{m,t}$ is the excess return on the market portfolio in period t, the coefficient C_p' captures the response of the manager's β to the public information, and the coefficient γ_{tmc} measures the sensitivity of β to the manager's private macroforecast.

We follow the procedure of Ferson and Harvey (1999) and construct conditional timing models, as in Eq. (6). Our lagged instrumental variables, Z_t, include (1) the 1-month lagged return differential between 3-month and 1-month Treasury bills; (2) the Standard and Poor's 500 dividend yield; (3) the default spread measured as the difference between the yields on Moody's Baa and Aaa corporate bonds; (4) the term spread as measured by the yield differences between 10-year and 1-year Treasury bonds; (5) and the lagged 1-month Treasury bill yield.

4.1.3. Multifactor performance evaluation

Fama and French (1993) report that spread variables that control for size and book-to-market value are explanatory variables in performance studies. Therefore, we conduct an additional robustness test that utilizes the extended three-factor market-timing model of Fama and French, which is:

$$R_{p,t} = \alpha_p + b_p R_{m,t} + s_p(\mathrm{SMB}_t) + h_p(\mathrm{HML}_t) + \gamma[R_{m,t}]^2 + \varepsilon_{p,t} \tag{7}$$

where SMB and HML are the small minus big and high minus low factors that capture the size and book-to-market factors in returns.

The FF-3 model is controversial. Ferson and Harvey (1999) categorize proponents and detractors of the FF-3 model into three camps. One group of proponents believes that firm-specific factors can be used to identify mispriced securities, while the other group of proponents argues that the factors are priced risk factors. Detractors argue that the factors can result from data snooping and/or biased data. Furthermore, Ferson et al. (1999) show that the construction of spread factors can cause them to appear as risk factors in a CAPM framework despite bearing no relation to cross-sectional risks.

4.1.4. Conditional multifactor performance evaluation

Ferson and Harvey (1999) reexamine the FF-3 model using a model conditioned on public information and Carhart (1997) extends the FF-3 model by including a momentum factor (PR1YR). Therefore, we combine those models and produce a conditional variant of Carhart's four factor timing model, which is

$$\begin{aligned} R_{p,t} = {} & \alpha_p + b_p R_{m,t} + s_p(\mathrm{SMB}_t) + h_p(\mathrm{HML}_t) + p_p(\mathrm{PR1YR}_t) + C_p'(Z_t R_{m,t}) \\ & + \gamma_{tmc}[R_{m,t}]^2 + \varepsilon_{p,t}. \end{aligned} \tag{8}$$

4.1.5. Relative performance evaluation

For a team to outperform an individual, the team does not have to beat the market, only the individual. This requires a direct comparison of timing and stock selection abilities. Therefore, we amend the models by examining the loadings of the differences between the excess returns of the equally weighted indexes of team-managed and individually managed funds for each investment objective. We perform this test for each of the models and indexes to ensure that our results are robust.

4.1.6. Individual level fund analysis

A shortcoming of merely examining indexes of funds to draw conclusions about manager ability is that outliers could potentially wield inordinate influence on results if the performance of a few funds varied appreciably from the average. This is not simply a statistical concern since Bettenhausen (1991) finds group judgments to be more extreme than the judgments made by individuals. Moreover, the propensity to shift positions may be enhanced in the group setting and result in conformity or herding behavior (Hartman and Nelson, 1996; Olsen, 1996; Bikhchandani et al., 1998). If extreme or herding behavior exists, it may affect the distributions of timing and selectivity coefficients as well as the Sharpe measures. Therefore, we analyze each fund and then examine the distributions of the timing and selectivity coefficients and Sharpe measures.

For these timing and selectivity tests, we utilize the HM model and CRSP VW index as the market proxy to estimate each fund's timing and selectivity coefficients. Prior to further statistical testing, we examine the Sharpe measures and the timing and selectivity coefficients for normality to ensure our selection of statistical tests are appropriate. We partition the data several ways to examine normality. First, we examine each management structure across all investment objectives and then we examine each investment objective by management structure.

For our test, we modified the conditional FF-3 model by adding an instrumental variable. MGMT = 1 for funds managed by an individual (and zero otherwise) to permit scrutiny of loading on the management factor. If management type is not important, factor loadings should be small and not statistically significant. Our model is:

$$R_{p,t} = \alpha_p + b_p R_{m,t} + s_p(\mathrm{SMB}_t) + h_p(\mathrm{HML}_t) + C_p'(Z_t r_{m,t}) + \gamma_{tmc}[R_{m,t}]^2 + m_p(\mathrm{MGMT}_t) + \varepsilon_{p,t}. \quad (9)$$

4.2. Empirical results

4.2.1. Results of unconditional methods of performance evaluation

Table 5 reports the empirical results of using Jensen's α as a performance measure with the CRSP VW index as a market proxy. We compute all standard errors and t-statistics using the Newey and West (1987) heteroskedasticity consistent covariance method to correct for heteroskedasticity and potential autocorrelation.[9] Panels A and B present the results for team-managed funds and funds managed by an individual, respectively. The coefficients of determination, R^2, for the team-managed and individually managed large value funds are 0.954 and 0.968, respectively, suggesting that the model is a good fit. The risk-adjusted performance, α, for both team-managed and individually managed funds in the large value classification is slightly negative (−0.72% and −0.00% annually, respectively) but not statistically significant. We conclude that there is no evidence that either management structure outperformed the index.

Large blend funds exhibit patterns that are similar to the patterns exhibited by the large value funds. The coefficients of determination remain high and the α's are slightly negative (but not statistically significant). Again, we find no evidence that either management structure outperformed the index.

Midcap fund risk-adjusted performance results are similar to the large blend and large value fund results in that both management structures produced small negative α's that are not statistically significant. However, the coefficients of determination for the midcap funds are appreciably lower than the other classifications (approximately 0.45) suggesting that the securities

[9] Lag length selection uses the procedure in Newey and West (1987).

Table 5
Market-model results using the CRSP VW index as a market proxy

Investment objective	N	R^2	α	β
Panel A: Team managed				
Large value	20	0.954	−0.0006 (0.97)	0.887*** (49.70)
Large blend	14	0.927	−0.0008 (0.63)	1.061*** (32.09)
Midcap	9	0.445	−0.0011 (0.20)	1.212*** (6.55)
Panel B: Managed by an individual				
Large value	58	0.968	−0.0000 (0.09)	0.934*** (56.26)
Large blend	57	0.908	−0.0007 (0.53)	1.031*** (28.57)
Midcap	25	0.448	−0.0020 (0.39)	1.213*** (6.08)
Panel C: Team–individual $R_{team_{o,t}} - R_{ind_{o,t}} = \alpha_{team} + \beta_{team}(R_{m,t}) + \varepsilon_t$				
Large value	58	0.156	−0.0006 (1.43)	0.047*** (4.81)
Large blend	57	0.052	−0.0001 (0.28)	0.030** (2.12)
Midcap	25	0.000	0.0010 (1.17)	−0.001 (0.04)

The sample period is from January 1, 1992 through December 31, 2001. The model is

$$R_{p,t} = \alpha_p + \beta_p(R_{m,t}) + \varepsilon_{p,t}$$

where $R_{p,t}$ is the excess return of the portfolio and $R_{m,t}$ is the market risk premium in period t. We use monthly returns of 90-day T-bills and the CRSP VW index to proxy for the risk-free and market returns, respectively. N is the number of funds utilized to fabricate the index. The absolute t-statistics, in parentheses, use the Newey and West (1987) heteroskedasticity consistent covariance method.
*, **, *** denote significant differences at the 10%, 5% and 1% levels respectively.

that these funds invest in are not similar to those of the benchmark index. This suggests that it may be prudent to repeat these performance tests using another index to ascertain whether different results are obtained when the market proxy is changed. We will further address this issue in the following section.

Panel C reports the results of a direct test of performance that uses the excess returns on the team-managed funds minus the excess returns on the individually managed funds as the dependent variable. In this formulation, α represents the differential excess risk-adjusted return of team-managed funds above that of their individually managed counterparts. In two of three cases, funds managed by an individual manager had higher net-of-fee excess returns. In no case was α statistically different from zero; therefore, we conclude that there is no significant performance difference. However, results in Table 4 showed that expenses for funds managed by individuals were 36 bp higher and that 12-b-1 fees were about 10 bp higher. Thus, expenses net of 12-b-1 fees are about 26 bp higher for individuals. Funds managed by an individual manager also have higher average load fees. According to Berk and Green's model (2004), investors compete with their fund flows to equate the ex ante net-of-fee returns across mutual funds. In a world where skilled managers earn rents, more-skilled managers may earn larger before-fee returns and capture larger fees. Thus, our finding of similar net-of-fee performance could be consistent with a world in which one type of structure delivers higher risk-adjusted gross returns but earns higher compensation. Thus, to the extent that the higher expense ratios of individually managed funds reflect higher manager compensation, the evidence is consistent with the view that individuals perform more skillfully than teams.[10]

[10] We thank Wayne Ferson for bringing this point to our attention.

Table 6 reports results of using the Henrikson-Merton (HM) model with the CRSP VW index as a market proxy. Panels A and B present the performance results for three classifications of funds with two management structures. For team-managed funds, large value and large blend classifications exhibit small negative α's, and midcap funds exhibit small positive α's. All classifications of funds managed by an individual exhibit small positive α's. However, consistent with much of the prior literature on mutual fund performance, the α's are not statistically different from zero in any investment classification or management structure combination. This finding is consistent with an efficient market where information and trading have costs that diminish net returns and with ability that captures rents.

As for market timing ability, the gammas (γ) of large value funds that are managed by a team are small and positive, but not statistically significant. Large blend funds that are managed by a team and both large value and large blend funds that are managed by an individual manager exhibit small negative γ's, but they are not statistically significant. Therefore, we conclude that there is no evidence of any timing ability for large value or large blend funds. However, midcap funds exhibit negative timing ability irrespective of the management structure. Furthermore, this negative timing is statistically significant at the 0.10 and 0.05 level for team-managed and individually managed funds, respectively.[11]

Despite indications of perverse timing, this negative outcome may not be the result of decisions made by portfolio managers. Ferson and Warther (1996) argue that some indications of perverse timing are the result of fund flows into funds that are correlated with expectations of future market performance. If managers cannot or do not rapidly invest these funds, the cash position of the fund will increase causing the β to decrease. Therefore, while it would appear that the manager decreased β at the wrong time, the β decrease was not due to a managerial decision but to an externality. Edelen (1999) finds that monthly cash flows are capable of explaining negative timing ability.[12] Panel C presents the relative selectivity and timing ability. None of the α's are statistically different from zero, suggesting that no security selection differences exist between the two competing management structures. As for γ's, only large blend funds exhibit non-zero differential market timing ability. Because γ is positive, it may suggest that teams that manage large blend funds make better timing decisions than individuals do. Note, however, that their α's are slightly more negative. Glosten and Jagannathan (1994) show that this could result from managers buying convexity through option-like trading.

4.2.2. Impact of selected market proxy

Another issue that demands investigation is the impact of the selected market proxy on the results. Therefore, we replicate the empirical tests using the CRSP EW index and S&P 500 index as market proxies. Our motive for using the CRSP EW index is that Kothari and Warner (1997) find that it mitigates some size-related bias, and produces the least model misspecification. Selection of the S&P 500 index is motivated by its use in other studies and its ready availability as a performance benchmark.

While the estimated metrics differ between indexes, our qualitative conclusion regarding the relative performance of the two management structures does not. There is no evidence that one

[11] One explanation for this is that the result is spurious (e.g., Jagannathan and Korajczyk, 1986). Therefore, we examine the γ's for the S&P Midcap 400, Russell Midcap, and Dow Jones US Midcap indices. γ's for all three indices were small and negative, but not statistically different from zero.

[12] Our interest here is not whether fund managers exhibit timing ability but rather whether fund managers exhibit differential timing ability. Both management structures must contend with redeploying any fund flows. The rapidity with which funds can be redeployed may differ between management structures, causing differential timing ability to be exhibited.

Table 6
Henriksson-Merton model results using the CRSP VW index as a market proxy

Investment objective	N	R^2	α	β	γ
Panel A: Team managed					
Large value	20	0.954	−0.0010 (0.87)	0.876*** (26.43)	0.022 (0.40)
Large blend	14	0.927	−0.0002 (0.18)	1.081*** (16.67)	−0.041 (0.50)
Midcap	9	0.452	0.0062 (0.89)	1.456*** (8.04)	−0.498* (1.97)
Panel B: Managed by an individual					
Large value	58	0.968	0.0000 (0.01)	0.936*** (29.70)	−0.005 (0.09)
Large blend	57	0.909	0.0010 (0.60)	1.087*** (18.38)	−0.115 (1.50)
Midcap	25	0.459	0.0070 (1.00)	1.519*** (7.49)	−0.623** (2.21)
Panel C: Team–individual $R_{team_{\alpha,t}} - R_{ind_{\alpha,t}} = \alpha_{team} + \beta_{team}(R_{m,t}) + \gamma_{team}(R_{m,t})D_t + \varepsilon_t$					
Large value	58	0.161	−0.0010 (1.66)	−0.060*** (2.51)	0.027 (0.71)
Large blend	57	0.082	−0.0012 (1.59)	−0.006 (0.27)	0.074* (1.70)
Midcap	25	0.018	−0.0008 (0.48)	−0.062 (0.81)	0.124 (0.98)

The sample period is from January 1, 1992 through December 31, 2001. The model is

$$R_{p,t} = \alpha_p + \beta_p(R_{m,t}) + \gamma_p(R_{m,t})D_t + \varepsilon_{p,t}$$

where $R_{p,t}$ is the excess return of the portfolio and $R_{m,t}$ is the market risk premium for each time period t. γ_p is the estimated timing measure ($D_t = 1$ if R_m is greater than R_f). We use monthly returns of 90-day T-bills and the CRSP VW index to proxy for the risk-free and market returns, respectively. N is the number of funds utilized to fabricate the index. The absolute t-statistics, in parentheses, use the Newey and West (1987) heteroskedasticity consistent covariance method.
*, **, *** denote significant differences at the 10%, 5% and 1% levels, respectively.

management structure is superior to the other structure. In summary, our results regarding the impact of management structure on relative performance are consistent with those reported earlier, suggesting that our results are not driven by benchmark selection.[13]

4.2.3. Robustness of unconditional methodological models

A lingering possibility is that timing differences exist between funds in the two management structures but the HM test is unable to expose them. Therefore, a crucial question is whether our results are robust with respect to the selected timing method. As a verification of the robustness of our results, we duplicate statistical analysis using the TM model.[14]

Results for the TM model using the CRSP VW index as a market proxy are similar to those using the HM measure. Of the three investment objectives studied, none produced α's that were statistically different from zero. However, the change from the HM model to the TM model caused a general decrease in α's for those funds. As for timing, γ's, none of the investment objectives exhibit statistically significant timing with the TM measure.

In summary, the patterns observed in our previous tests are qualitatively unchanged. The α's are not statistically different from zero.[15] Timing measures remain negative, but generally not statistically different from zero.

[13] Results using the Jensen and HM models with alternate indexes are available from the corresponding author.
[14] For brevity, only results for the CRSP VW index are presented here. Robustness tests using the S&P 500, CRSP VW, and CRSP EW indexes are available from the corresponding author. However, those results fail to detect differential performance between the two competing management structures.
[15] Only the α's of large value funds using the CRSP EW index as a benchmark are statistically non-zero. Furthermore, those α's are positive for both individually managed and team-managed funds.

Table 7
Conditional Treynor-Mazuy model results using the CRSP VW index

Investment objective	N	R^2	α	b	γ
Panel A: Team managed					
Large value	20	0.959	−0.0180** (2.35)	0.892*** (42.79)	0.276 (0.83)
Large blend	14	0.933	0.0317** (1.99)	1.063*** (29.21)	−0.093 (0.23)
Midcap	9	0.492	0.1231 (1.40)	1.193*** (5.97)	−3.456 (1.32)
Panel B: Managed by an individual					
Large value	58	0.971	−0.0204*** (2.91)	0.939*** (52.12)	0.258 (0.93)
Large blend	57	0.916	0.0359* (1.73)	1.029*** (24.68)	−0.544 (0.93)
Midcap	25	0.494	0.1148 (1.34)	1.195*** (5.47)	−3.463 (1.15)
Panel C: Team–individual					
Large value	58	0.211	0.0024 (0.38)	−0.047*** (4.36)	0.018 (0.11)
Large blend	57	0.150	−0.0042 (0.60)	0.034** (2.46)	0.452* (1.97)
Midcap	25	0.058	0.0083 (0.84)	−0.002 (0.05)	0.007 (0.01)

The sample period is from January 1, 1992 through December 31, 2001. The model is

$$R_{p,t} = \alpha_p + b_p R_{m,t} + C_p'(Z_t R_{m,t}) + \gamma_{tmc}[R_{m,t}] + \varepsilon_{p,t}$$

where $R_{p,t}$ is the excess return of the portfolio and $R_{m,t}$ is the excess return on the market proxy in period t. Our lagged instrumental variables, Z_t, include (1) the 1-month lagged return differential between 3-month and 1-month Treasury bills; (2) the Standard and Poor's 500 dividend yield; (3) the default spread measured as the difference between the yields on Moody's Baa and Aaa corporate bonds; (4) the term spread as measured by the yield differences between 10-year and 1-year Treasury bonds; and (5) the lagged 1-month Treasury bill yield. We use monthly returns of 90-day T-bills and the CRSP VW index to proxy for the risk-free and market returns, respectively. N is the number of funds utilized to fabricate the index. The absolute t-statistics, in parentheses, use the Newey and West (1987) heteroskedasticity consistent covariance method.
*, **, *** denote significant differences at the 10%, 5% and 1% levels, respectively.

4.2.4. Results of conditional performance evaluation models

Results of the conditional Treynor-Mazuy model in Table 7 illustrate that conditioning marginally improved the coefficients of determination and enlarged the absolute magnitude of the t-statistics on the selectivity coefficients. With the conditional models, large value funds exhibit negative selectivity and large blend funds exhibit positive selectivity. These results are independent of management structure and are statistically significant at the 0.10 level or better. As for timing, none of the investment objective and management structure combinations exhibit statistical significance. However, Panel C shows that while no statistically significant selectivity differences exist for any of the investment objectives, large blend funds managed by a team exhibit positive differential timing ability that is marginally significant. However, their α's are slightly more negative. Again, these results may suggest managers are procuring convexity through option-like trading (e.g., Glosten and Jagannathan, 1994).

4.2.5. Results of multifactor performance evaluation models

Results of the FF-3 timing models, in Table 8, reveal that the coefficients of determination are larger than with previous models. All coefficients on the HML variables are statistically significant and four out of six coefficients on the SMB variables are significant. All fund class and management structure combinations exhibit negative selection measures, α's, but none of them is statistically significant. As for timing, all fund class and management structure combinations exhibit positive timing coefficients, γ's, but none of them is statistically significant. Because none

Table 8
Fama and French three-factor Treynor-Mazuy regressions using the CRSP VW index

Investment objective	N	R^2	α	b	s_p	h_p	γ
Panel A: Team managed							
Large value	20	0.968	−0.0012 (1.64)	0.911*** (52.88)	0.033 (1.12)	0.095*** (4.00)	0.041 (0.18)
Large blend	14	0.959	−0.0006 (0.59)	1.031*** (32.76)	0.022 (0.65)	−0.147*** (6.53)	0.249 (0.86)
Midcap	9	0.901	−0.0006 (0.27)	1.108*** (14.47)	0.671*** (12.75)	−0.589*** (8.29)	0.492 (0.81)
Panel B: Managed by an individual							
Large value	58	0.980	−0.0008 (1.52)	0.960*** (68.73)	0.058*** (3.54)	0.096*** (7.54)	0.147 (0.88)
Large blend	57	0.967	−0.0003 (0.35)	0.997*** (33.58)	0.085** (2.55)	−0.169*** (6.84)	0.108 (0.36)
Midcap	25	0.921	−0.0018 (0.72)	1.116*** (12.03)	0.719*** (12.27)	−0.569*** (9.94)	0.519 (0.52)
Panel C: Team–individual							
Large value	58	0.198	−0.0004 (0.94)	−0.049*** (4.24)	−0.025 (1.28)	−0.001 (0.08)	−0.106 (0.77)
Large blend	57	0.459	−0.0030 (0.64)	0.03*** (3.87)	−0.062*** (5.85)	0.022*** (2.75)	0.141 (1.26)
Midcap	25	0.024	0.0012 (1.13)	−0.008 (0.23)	−0.047 (1.15)	−0.021 (0.58)	−0.027 (0.05)

The sample period is from January 1, 1992 through December 31, 2001. The model is

$$R_{p,t} = \alpha_p + b_p R_{m,t} + s_p(\mathrm{SMB}_t) + h_p(\mathrm{HML}_t) + \gamma_p [R_{m,t}]^2 + \varepsilon_{p,t}$$

where $R_{p,t}$ is the excess return of the portfolio and $R_{m,t}$ is the excess return on the market proxy in period t. γ_p is the estimated timing coefficient. We use monthly returns of 90-day T-bills and the CRSP VW index to proxy for the risk-free and market returns, respectively. N is the number of funds utilized to fabricate the index. The absolute t-statistics, in parentheses, use the Newey and West (1987) heteroskedasticity consistent covariance method.
*, **, *** denote significant differences at the 10%, 5% and 1% levels, respectively.

of the fund classes exhibit differential timing or selectivity ability (in Panel C), we conclude that there is no difference in performance due to management structure.

4.2.6. Results of conditional multifactor performance evaluation models

Table 9 presents results of combining the FF-3 and conditional models and adding Carhart's (1997) momentum variable. The coefficients of determination are the highest of any model that we have utilized thus far (with poorest fit explaining nearly 91% of the change in the dependent variables returns). This finding is not surprising because we have combined the predictor variables. The coefficients on the HML and SMB remain significant, but the coefficient on the momentum variable is only statistically significant for midcap funds managed by a team. Therefore, our results suggest that the momentum variable has little explanatory power for our sample of funds. Our primary interest, however, centers on selectivity and timing differences between the two competing management structures. As for selectivity, there is a mix of positive and negative coefficients, but the signs are the same for each investment objective class, suggesting that management structure is not a vital determinant of security selection. In only one case was the selectivity measure statistically significant. We find that large value funds managed by an individual have negative selection ability. Again, the α's and γ's for these funds go in opposite directions. Timing coefficients are positive, but not statistically significant, for all investment objective and management structure combinations.

Our primary focus of differential performance is shown in Panel C. None of the differential α's or γ's is statistically different from zero. Therefore, our initial conclusions remain unaltered.

4.2.7. Survivorship and incubation bias free sample results

As a final robustness test of macro level results, we cull a new sample of funds (living and dead) in the large growth, aggressive growth, and growth and income categories from the CRSP Survivorship Bias Free database. The initial sample provides 48 large growth (23 team/25 individual), 40 aggressive growth (15 team/25 individual), and 40 growth and income (25 team/15 individual) funds. No fund in this sample is an index fund and none of the funds list a change in management structure in the CRSP database. To mitigate incubation bias issues, discussed in Section 3.2, we eliminate the first 48 months of a new fund's returns. Creating the incubation bias free sample resulted in eliminating several funds. The final sample consists of 46 large growth (23 team/23 individual), 36 aggressive growth (15 team/21 individual), and 36 growth and income (24 team/12 individual) funds.

Our primary interest remains the selectivity and timing coefficients for the two competing management structures. Results of combining the FF-3 and conditional models and adding a momentum variable show that there is a mix of positive and negative selectivity coefficients, but again none of them are statistically significant.[16] As for timing, large growth and aggressive growth funds have positive coefficients and growth and income funds have negative coefficients. For both management structures, the timing coefficients for aggressive growth funds are statistically significant, whereas the coefficients for growth and income funds are not statistically significant. Again, the α's and γ's have opposite signs. As for differential performance, none of the α's or γ's are statistically different from zero. Therefore, our initial conclusions remain unaltered.

4.2.8. Individual level fund performance metric distributions

Fig. 1 portrays the three distributions of interest for the entire sample of funds partitioned by management structure. The general shape of the distributions of α's for the individually managed

[16] Complete results are available from the corresponding author.

Table 9
Conditional Carhart four-factor Treynor-Mazuy model using the CRSP VW index

Investment objective	N	R^2	α	b	s_p	h_p	p_p	γ
Panel A: Team managed								
Large value	20	0.971	−0.0072 (0.96)	0.909*** (63.69)	0.034 (1.64)	0.086*** (4.46)	−0.009 (0.37)	0.170 (0.52)
Large blend	14	0.965	0.0150 (1.57)	0.991*** (37.68)	−0.011 (0.47)	−0.217*** (5.87)	−0.084 (1.92)	0.129 (0.49)
Midcap	9	0.908	−0.0078 (0.43)	1.192*** (12.93)	0.723*** (10.31)	−0.466*** (5.02)	0.146* (1.82)	1.075 (1.49)
Panel B: Managed by an individual								
Large value	58	0.982	−0.0116** (2.17)	0.961*** (85.11)	0.058*** (4.37)	0.091*** (4.77)	−0.001 (0.06)	0.262 (1.12)
Large blend	57	0.971	0.0107 (1.03)	0.971*** (34.36)	0.059** (2.19)	−0.216*** (6.55)	−0.055 (1.40)	0.052 (0.18)
Midcap	25	0.922	−0.0101 (0.55)	1.126*** (11.50)	0.712*** (9.56)	−0.573*** (6.64)	0.008 (0.10)	0.842 (0.84)
Panel C: Team–individual								
Large value	58	0.249	0.0044 (0.82)	−0.052*** (3.58)	−0.024 (1.43)	−0.005 (0.36)	−0.007 (0.63)	−0.092 (0.56)
Large blend	57	0.518	0.0040 (1.10)	0.020 (1.77)	−0.071*** (5.37)	−0.001 (0.09)	−0.029*** (3.01)	0.077 (0.61)
Midcap	25	0.266	0.0023 (0.23)	0.066** (2.43)	0.010 (0.34)	0.106*** (3.05)	0.138*** (5.13)	0.232 (0.55)

The sample period is from January 1, 1992 through December 31, 2001. The model is

$$R_{p,t} = \alpha_p + b_p R_{m,t} + s_p(\mathrm{SMB}_t) + h_p(\mathrm{HML}_t) + p_p(\mathrm{PR1YR}_t) + C_p'(Z_t R_{m,t}) + \gamma_{tmc}[R_{m,t}]^2 + \varepsilon_{p,t}$$

where $R_{p,t}$ is the excess return of the portfolio and $R_{m,t}$ is the excess return on the market proxy in period t. Our lagged instrumental variables, Z_t, include (1) the 1-month lagged return differential between 3-month and 1-month Treasury bills; (2) the Standard and Poor's 500 dividend yield; (3) the default spread measured as the difference between the yields on Moody's Baa and Aaa corporate bonds; (4) the term spread as measured by the yield differences between 10-year and 1-year Treasury bonds; and (5) the lagged 1-month Treasury bill yield. We use monthly returns of 90-day T-bills and the CRSP VW index to proxy for the risk-free and market returns, respectively. N is the number of funds utilized to fabricate the index. The absolute t-statistics, in parentheses, use the Newey and West (1987) heteroskedasticity consistent covariance method.
*, **, *** denote significant differences at the 10%, 5% and 1% levels, respectively.

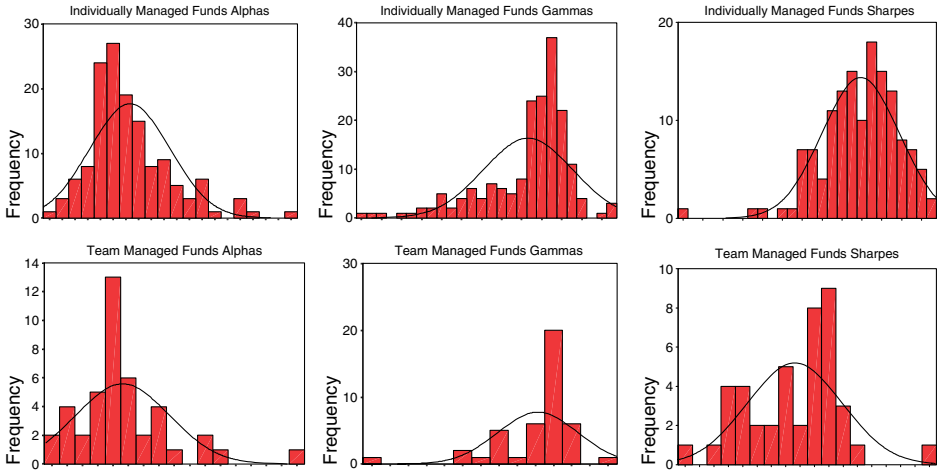

Fig. 1. Individually managed and team-managed funds α's, γ's, and Sharpe distributions. The distributions of α's, γ's, and Sharpe measures for the entire sample of funds during our sample period from January 1, 1992 through December 31, 2001 are presented. The top row provides distributions for funds managed by an individual and the bottom row provides distributions for funds managed by managed by a team.

fund subsample and the team-managed fund subsample is similar in that both exhibit positive skewness. To test whether each distribution is normally distributed, we use the Jarque-Bera statistic.

The Jarque-Bera test results in Table 10 for each investment-objective management-structure combination suggest that the distributions of α's, γ's, and Sharpe measures for midcap funds are Gaussian with the exception of the Sharpe measure of individually managed funds. For large blend funds, all coefficients except the α's for funds managed by an individual are Gaussian. The distributions of α's, γ's, and Sharpe measures for large value are Gaussian with the exception of the γ's of funds managed by an individual. This positive skewness may suggest that previous tests overlooked the existence of some high-performance individuals, which strengthens our finding that teams do not perform better.

4.2.9. Individual level fund performance analysis

To ascertain relative performance, we compare the performance measures of funds managed by an individual manager to their counterparts managed by a team. For distributions that are Gaussian, we use a *t*-test for hypothesis testing. To conduct hypothesis testing for the non-Gaussian distributions, we use the Mann-Whitney *U* test since it is more powerful than the median test. The Mann-Whitney test examines whether two sampled populations are equivalent in location. The observations from both groups are combined and ranked (the average rank is assigned for ties). If the populations are similar, the ranks should be equally mixed between the two samples.

Mann-Whitney tests are used to examine the distributions of (1) α's, γ's, and Sharpe measures for the total sample of funds, (2) midcap Sharpe, (3) large blend α's and γ's, and (4) large value γ's. Of these seven tests, only the γ's of the combined sample are statistically different, and that difference is marginally significant at the 0.10 level. These tests further support the classical decision-making theory.[17]

[17] Complete results for all individual level fund analysis are available from the corresponding author.

Table 10
Coefficient distributions of individual funds

Sample	α					γ					Sharpe					N
	Mean	Median	Skew	Kurt	Jarque-Bera	Mean	Median	Skew	Kurt	Jarque-Bera	Mean	Median	Skew	Kurt	Jarque-Bera	
Full	0.001	0.001	1.24	5.26	0.00	−0.148	−0.067	−1.33	4.88	0.00	0.135	0.141	−0.73	4.83	0.00	183
Large blend (all)	0.001	0.000	0.75	3.77	0.01	−0.100	−0.069	−1.09	4.59	0.00	0.141	0.149	−0.59	4.04	0.03	71
Large value (all)	−0.000	0.000	−0.07	3.04	0.97	0.002	0.010	0.21	4.80	0.00	0.151	0.156	−0.44	3.25	0.25	78
Midcap (all)	0.007	0.006	0.80	3.08	0.16	−0.590	−0.552	−0.36	3.10	0.69	0.086	0.087	−1.79	8.42	0.00	34
Individual (all)	0.002	0.001	1.17	4.96	0.00	−0.160	−0.079	−1.20	4.36	0.00	0.136	0.141	−0.86	4.98	0.00	140
Team (all)	0.001	0.000	1.55	6.79	0.00	−0.107	−0.018	−1.81	7.14	0.00	0.133	0.144	−0.06	3.43	0.78	43
Midcap (I)	0.007	0.007	0.67	2.94	0.39	−0.622	−0.573	−0.36	2.93	0.76	0.082	0.082	−1.73	7.19	0.00	25
Midcap (T)	0.006	0.005	1.21	3.81	0.30	−0.498	−0.415	−0.83	3.41	0.58	0.100	0.092	0.84	3.05	0.59	9
Large blend (I)	0.001	0.000	0.79	3.42	0.04	−0.115	−0.082	−1.03	4.24	0.00	0.140	0.147	−0.50	3.75	0.16	57
Large blend (T)	−0.000	0.000	−0.40	3.59	0.75	−0.041	−0.019	−0.65	4.02	0.45	0.145	0.155	−1.15	3.22	0.21	14
Large value (I)	0.000	0.000	−0.13	3.04	0.92	−0.005	−0.003	0.50	5.00	0.00	0.155	0.159	−0.48	2.83	0.31	58
Large value (T)	−0.001	−0.000	−0.39	2.30	0.64	0.022	0.041	−0.60	4.95	0.11	0.139	0.152	−0.17	3.63	0.81	20

Descriptive statistics for the estimated timing, selectivity, and Sharpe measures for our sample of funds during the sample period from January 1, 1992 through December 31, 2001 are presented. Measures are computed for each fund and then grouped to examine the behavior of the sample groups of interest. Column 1 lists the sample groups of interest. Columns 2 through 6 present the mean, median, skewness, kurtosis, and Jarque-Bera (p-value) test for normality of the selectivity coefficient. Columns 7 through 11 and 12 through 16 report the mean, median, skewness, kurtosis, and Jarque-Bera (p-value) test for normality of the timing coefficients and Sharpe measures, respectively. The sample size is listed in column 17.

For the Gaussian coefficient distributions, we employ the *t*-test to learn whether differences between individually managed and team-managed funds exist. To learn whether to employ two-tailed homoskedastic or heteroskedastic *t*-tests, we utilized the Levene test to determine if variances are equal. Results suggest that the homoskedastic *t*-test is the appropriate test for the distributions of midcap α's and γ's, large blend Sharpes, and large value α's and Sharpes. Homoskedastic *t*-test results suggest that no difference exists between any of the Gaussian coefficient distributions of α's, γ's, or Sharpe measures for the two competing management structures, providing additional evidence in favor of the classical decision making theory.

For our subsequent test, we create Bernoulli random variables for the α's and γ's of funds. Positive (negative) values of the estimated coefficients are coded as successes (failures) and we examine whether the proportion of positive and negative timing and selectivity coefficients is different from the expected under the null provided by the EMH. The chi-square test results shown in Table 11 suggest that the cross-sectional examination of all investment objectives and management structures reject the nulls of no selectivity and/or timing ability. The entire sample produces more positive selectivity coefficients and fewer positive timing coefficients than expected under the null. Similar results are revealed for the entire cross section of funds managed by an individual. Only large blend funds managed by an individual exhibit perverse timing and only large value funds managed by a team exhibit superior selectivity.

For our concluding test, we modified the conditional FF-3 model by adding an instrumental variable, MGMT, that takes on a value of one for funds managed by an individual (and zero otherwise) to permit scrutiny of loading on the management factor. Results suggest that the management structure is not important for large blend funds, midcap funds, or for the combined sample of all investment objectives. However, the loading on the MGMT instrumental variable for large value funds was a small positive which was statistically significant ($p<0.04$). This is further support that teams do not make superior investment decisions.

Table 11
Chi-square coefficient tests

Performance measure	Investment class	Management structure	*N*	Mean	Standard deviation	Fraction positive	*p*-value	Fraction team
Alpha	All	All	183	0.612	0.489	112/183	0.002	43/183
Gamma	All	All	183	0.311	0.464	57/183	0.000	43/183
Alpha	All	Individual	140	0.636	0.483	89/140	0.001	0/140
Alpha	All	Team	43	0.535	0.505	23/43	0.647	43/43
Gamma	All	Individual	140	0.288	0.453	40/140	0.000	0/140
Gamma	All	Team	43	0.395	0.495	17/43	0.170	43/43
Alpha	Large blend	Individual	57	0.544	0.502	31/57	0.508	0/57
Alpha	Large blend	Team	14	0.571	0.514	8/14	0.593	14/14
Gamma	Large blend	Individual	57	0.211	0.411	12/57	0.000	0/57
Gamma	Large blend	Team	14	0.286	0.469	4/14	0.109	14/14
Alpha	Large value	Individual	58	0.569	0.500	33/58	0.294	0/58
Alpha	Large value	Team	20	0.300	0.470	6/20	0.074	20/20
Gamma	Large value	Individual	58	0.483	0.504	28/58	0.793	0/20
Gamma	Large value	Team	20	0.650	0.489	13/20	0.180	20/20

Chi-square test results of whether the proportion of positive and negative timing and selectivity coefficients are different from the expected under the null provided by the EMH for our sample of funds during the sample period from January 1, 1992 through December 31, 2001 are presented. Columns 1 through 4 list the performance measure, *Morningstar* investment objective, management structure, and sample size. The mean, standard deviation, fraction positive, *p*-value of the chi-square test, and fraction team managed are presented in columns 5 through 9, respectively.

5. Conclusion

Two opposing theories of decision making are the classical decision-making theory and the behavioral decision-making theory. From the classical perspective, whether an individual, group, or organization makes the decision, it should lead to the same maximizing choice and optimal performance outcome. However, the behavioral theory asserts that when the task is complex, and completed under high levels of uncertainty, group members tend to pool and integrate their resources, correcting each other's errors (e.g., Hinsz et al., 1997), producing qualitatively and quantitatively superior performance.

With trillions of dollars invested in open-end mutual funds, and thousands of funds in existence, the mutual fund industry is a prime candidate for use in testing decision-making theories. Therefore, we seek to learn whether teams or individuals make better market timing and security selection decisions on average. Our empirical tests comprise a multiplicity of parametric market timing methods. Despite extensive analyses using a multiplicity of models and benchmarks, we fail to find any evidence to support the behavioral decision-making theory.

Acknowledgement

The authors thank Wayne Ferson and two anonymous referees for insightful comments that aided in crafting the final paper, Bruce Lehmann for comments on an earlier draft of this paper, and Kenneth French who provided SMB, HML, and momentum factor data. Larry Prather thanks East Tennessee State University's Research and Development Committee for providing funding.

Appendix A. Stability of the quadratic term

Another meaningful issue pertains to the stability of the quadratic term since some researchers find that quadratic regressions produce estimates that change signs across subsamples, and sometimes even with the addition or deletion of several observations. Since the stability of these quadratic terms is important to our study, we conduct multiple Chow breakpoint tests to investigate the stability of the regression coefficients. Our testing encompasses each combination of an index, timing test (HM and TM), and four breakpoint dates (June 1994, January 1997, June 1999, and one test using all three dates). We selected 30-month and 60-month periods to examine coefficient stability. The Chow breakpoint test statistic is:

$$F = \frac{(\tilde{u}'\tilde{u} - (u_1'u_1 + \tilde{u}_2 u_2))/k}{(u_1'u_1 + u_2'u_2)/(T-2k)}, \tag{10}$$

where $\tilde{u}'\tilde{u}$ is the restricted sum of squared residuals, $u_i'u_i$ is the sum of squared residuals from subsample i, T is the number of observations, and k is the number of parameters in the equation.

Results of the 168 tests suggest that stability can be rejected 14 times at the 1% level, 26 times at the 5% level, and 22 times at the 10% level. These test results can be coded as successes (failures to reject the null hypothesis) and failures (rejections of the null hypothesis) to create Bernoulli trials which will form a dichotomous (binomial) distribution. Because the sample size is large and both np and $(1-np)$ are greater than five, the normal approximation of the binomial can be utilized to ascertain whether the observed rejections exceed the rejections that could be explained by chance. At the 10% level, we would expect 16.8 rejects by chance alone (Type I errors). However, we observed 62 rejections at the 10% level or higher (36.9%), which falls

outside a confidence interval based on random chance. This suggests that coefficient instability has some impact on our sample.

The pattern of rejections suggests that the majority of rejections occur when comparing short periods with longer periods. When comparing the two five-year periods with the entire sample period, only 11 rejections were unearthed out of 42 trials (26.2%). Again, this falls outside a confidence interval based on chance alone.

Because our rejection rates exceed those that can reasonably be expected due to chance alone, we acknowledge that coefficient instability exists to some degree. To ascertain the impact of this instability, we separate our sample at the midpoint and run both HM and TM models for each subsample using all four benchmarks. These results suggest that our conclusions are unaltered.

References

Association for Investment Management and Research, 1997. AIMR Performance Presentation Standards Handbook. Association for Investment Management and Research, Charlottesville.

Arteaga, K., Ciccotello, C., Grant, T., 1998. New equity funds: marketing and performance. Financial Analysts Journal 54, 43–50.

Barber, B., Odean, T., 1999. The courage of misguided convictions. Financial Analysts Journal 55, 41–55.

Becker, C., Ferson, W., Myers, D., Schill, M., 1999. Conditional market timing with benchmark investors. Journal of Financial Economics 52, 119–148.

Bello, Z., Janjigian, V., 1997. A reexamination of the market-timing and security-selection performance of mutual funds. Financial Analysts Journal 53, 24–30.

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

Bettenhausen, K., 1991. Five years of groups research: what we've learned and what needs to be addressed. Journal of Management 17, 345–382.

Bikhchandani, S., Hirshleifer, D., Welch, L., 1998. Learning from the behavior of others: conformity, fads, and informational cascades. Journal of Economic Perspectives 12, 151–170.

Bloomfield, R., Libby, R., Nelson, M., 2000. Underreactions, overreactions and moderated confidence. Journal of Financial Markets 3, 113–137.

Brown, S., Goetzmann, W., 1995. Performance persistence. Journal of Finance 50, 679–698.

Brown, S., Goetzmann, W., 1997. Mutual fund styles. Journal of Financial Economics 43, 373–399.

Brown, S., Goetzmann, W., Ibbotson, R., Ross, S., 1992. Survivorship bias in performance studies. Review of Financial Studies 5, 553–580.

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

Chevalier, J., Ellison, G., 1999a. Are some mutual fund managers better than others? Journal of Finance 54, 875–899.

Chevalier, J., Ellison, G., 1999b. Career concerns of mutual fund managers. Quarterly Journal of Economics 114, 389–432.

Christopherson, J., Ferson, W., Glassman, D., 1998. Conditioning manager alphas on economic information: another look at the persistence of performance. Review of Financial Studies 11, 111–142.

Connor, G., Korajczyk, R., 1991. The attributes, behavior, and performance of U.S. mutual funds. Review of Quantitative Finance and Accounting 1, 5–26.

Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under- and overreaction. Journal of Finance 53, 1839–1885.

Daniel, K., Titman, S., 1999. Market efficiency in an irrational world. Financial Analysts Journal 55, 28–40.

Edelen, R., 1999. Investor flows and the assessed performance of open-end mutual funds. Journal of Financial Economics 53, 439–456.

Evans, R. Does alpha really matter? Evidence from mutual fund incubation, termination, and manager change. Unpublished working paper, Carroll School of Management, Boston College; 2005.

Fama, E, French, K, 1993. Common risk factors in the returns on bonds and stocks. Journal of Financial Economics 33, 3–53.

Ferson, W., Harvey, C., 1999. Conditioning variables and the cross-section of stock returns. Journal of Finance 54, 1325–1360.

Ferson, W., Schadt, R., 1996. Measuring fund strategy and performance in changing economic conditions. Journal of Finance 51, 425–461.

Ferson, W., Warther, V., 1996. Evaluating fund performance in a dynamic market. Financial Analysts Journal 52, 20–28.

Ferson, W., Sarkissian, S., Simin, T., 1999. The alpha factor asset-pricing model: a parable. Journal of Financial Markets 2, 49–68.
Glosten, L., Jagannathan, R., 1994. A contingent claim approach to performance evaluation. Journal of Empirical Finance 1, 133–160.
Golec, J., 1996. The effects of mutual fund managers' characteristics on their portfolio performance, risk, and fees. Financial Services Review 5, 133–148.
Gottesman A, Morey M. Manager education and mutual fund performance. Journal of Empirical Finance; in press.
Hambrick, D., Mason, P., 1984. Upper echelons: the organization as a reflection of its top managers. Academy of Management Review 9, 193–206.
Hartman, S., Nelson, B., 1996. Group decision making in the negative domain. Group and Organization Management 21, 146–162.
Henriksson, R., Merton, R., 1981. On market timing and investment performance II: statistical procedures for evaluating forecasting skills. Journal of Business 54, 513–533.
Hinsz, V., Tindale, R., Vollrath, D., 1997. The emerging conceptualization of groups as information processors. Psychological Bulletin 121, 43–64.
Hogarth, R., Reder, M., 1987. Introduction: Perspectives From Economics and Psychology. In: Hogarth, R.M., Reder, M.W. (Eds.), Rational choice: the contrast between economics and psychology. University of Chicago Press, Chicago, pp. 1–23.
Hollenbeck, J., Ilgen, D., Sego, D., Hedlund, J., Major, D., Phillips, J., 1995. Multilevel theory of team decision making: decision performance in teams incorporating distributed expertise. Journal of Applied Psychology 80, 292–317.
Jagannathan, R., Korajczyk, R., 1986. Assessing the market timing performance of managed portfolios. Journal of Business 59, 217–235.
Jensen, M., 1968. The performance of mutual funds in the period 1945–1964. Journal of Finance 23, 389–419.
Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 4, 263–291.
Kahneman, D., Tversky, A., 1984. Choices, values, and frames. American Psychologist 39, 341–350.
Kameda, T., Davis, J., 1990. The function of the reference point in individual and group risk decision making. Organizational Behavior and Human Decision Processes 46, 55–76.
Klemkosky, R., 1976. Additional evidence on the risk level discriminatory powers of the Weisenberger classifications. Journal of Business 45, 48–50.
Kothari S, Warner J. Evaluating mutual fund performance. Unpublished working paper, Sloan School of Management, MIT; 1997.
Lehmann, B., Modest, D., 1987. Mutual fund performance evaluation: a comparison of benchmarks and benchmark comparisons. Journal of Finance 42, 233–265.
McNamara, G., Bromiley, P., 1997. Decision making in an organizational setting: cognitive and organizational influences on risk assessment in commercial lending. Academy of Management Journal 40, 1063–1088.
McShane, S., Von Glinow, M., 2000. Organizational behavior. Irwin McGraw-Hill, Boston.
Mellers, B., Schwartz, A., Cooke, A., 1998. Judgment and decision making. Annual Review of Psychology 49, 447–477.
Newey, W., West, K., 1987. A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Odean, T., 1998. Volume, volatility, price, and profit when all traders are above average. Journal of Finance 53, 1887–1934.
Olsen, R., 1996. Implications of herding behavior for earnings estimation, risk assessment, and stock returns. Financial Analysts Journal 52, 37–41.
Porter, G., Trifts, J., 1998. Performance persistence of experienced mutual fund managers. Financial Services Review 7, 57–68.
Pressman, S., 1998. On financial frauds and their causes: investor overconfidence. American Journal of Economics and Sociology 57, 405–421.
Sharpe, W., 1966. Mutual fund performance. Journal of Business 39, 119–138.
Shukla, R., Singh, S., 1994. Are CFA charterholders better equity fund managers? Financial Analysts Journal 50, 68–74.
Sniezek, J., Henry, R., 1989. Accuracy and confidence in group judgment. Organizational Behavior and Human Decision Processes 43, 1–28.
Thaler, R., 1999. The end of behavioral finance. Financial Analysts Journal 55, 12–17.
Treynor, J., Mazuy, K., 1966. Can mutual funds outguess the market? Harvard Business Review 44, 131–136.
Vollrath, D., Sheppard, B., Hinsz, V., Davis, J., 1989. Memory performance by decision-making groups and individuals. Organizational Behavior and Human Decision Processes 43, 289–300.
Zeckhauser, R., 1987. Comments: Behavioral Versus Rational Economics: What You See is What You Conquer. In: Hogarth, R.M., Reder, M.W. (Eds.), Rational choice: the contrast between economics and psychology. University of Chicago Press, Chicago, pp. 251–265.