NBER WORKING PAPER SERIES

HOT HANDS IN MUTUAL FUNDS: THE PERSISTENCE OF PERFORMANCE, 1974-87

Darryll Hendricks

Jayendu Patel

Richard Zeckhauser

Working Paper No. 3389

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 1990

The research was supported by grants from the Bradley Foundation and the Decision, Risk, and Management Science Program of the National Science Foundation. We thank Robert Zeckhauser, whose informal studies of the hot hands phenomenon stimulated this investigation, and Sheridan Titman for providing data on the P8 indices. Helpful comments from Richard Ruback, Larry Summers and seminar participants at the Harvard Business School, the Kennedy School of Government, and Cornell University are gratefully acknowledged. This paper is part of NBER's research program in Financial Markets and Monetary Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #3389
June 1990

HOT HANDS IN MUTUAL FUNDS: THE PERSISTENCE OF PERFORMANCE, 1974-87

ABSTRACT

The net returns of no-load mutual growth funds exhibit a hot-hands phenomenon during 1974-87. When performance is measured by Jensen's alpha, mutual funds that perform well in a one year evaluation period continue to generate superior performance in the following year. Underperformers also display short-run persistence. Hot hands persists in 1988 and 1989.

The success of the hot hands strategy does not derive from selecting superior funds over the sample period. The timing component -- knowing when to pick which fund -- is significant. These results are robust to alternative equity portfolio benchmarks, such as those that account for firm-size effects and mean reversion in returns. Capitalizing on the hot hands phenomenon, an investor could have generated a significant, risk-adjusted excess return of 10% per year.

Darryll Hendricks
Jayendu Patel
Richard Zeckhauser
John F. Kennedy School of Government
Harvard University
Cambridge, MA 02138

I. Introduction and Overview

Mutual fund performance has been extensively studied on behalf of investors seeking practical advice, and by academics testing the efficient markets hypothesis (EMH). Studies since the 1960s — see the classic papers by Treynor (1965), Sharpe (1966), and Jensen (1968), and recent updates with refinements by Shawky (1982), Ippolito (1989), Grinblatt and Titman (1989), and references therein — have forged an academic consensus that mutual funds do not offer ex-ante net returns to prospective investors that are superior to benchmark portfolios (like the 'market' portfolio), though gross returns of funds outperform passive strategies sufficiently to cover their fees and loads.

The practitioner literature sees matters differently. For instance, Rugg (1986) advocates, with some caveats, investing in aggressive-growth equity funds that are top-ranking performers in the most recent phase (one to six months) of a bull market. Similarly, *Consumer Guide* (1988) reports, "Loads, fees, and expenses can be considerable, but most financial professionals suggest that the performance of the fund, not the costs, should be the primary consideration when choosing a fund."[1] Mutual fund performance rankings are compiled on a regular and timely basis and are widely followed. Mutual funds that do relatively well tout their performance prominently in their advertising. Those that don't, search for the measure that puts them in the best possible light. Directly or indirectly, investors are willing to act on such ongoing relative performance information. In Patel, Zeckhauser, and Hendricks (1990), we document that investors steer their money to funds that have performed well recently. Are such investor behaviors justified?[2]

---

[1]In contrast, academics, as in Brealey's (1983) chapter on "Can Professional Investors Beat the Market?", advise that most of the differences between ex post performances of individual funds are due to chance.

[2]Numerous biases in individual decision-making have been identified. Tversky and Kahneman (1971) have shown that people expect the properties of large samples, such as convergence of relative frequency to population parameters, to hold in small samples too. From such evidence, it is easy to see why investors, who may have a proclivity to generalize too readily from small samples, may incorrectly infer autocorrelation in performance from observed runs in mutual fund returns that arise from random stock selection and random market timing.

We reexamine the quarterly performance data of open-end, no-load equity funds. We evaluate statistical evidence of the persistence of superior performance by identifiable mutual funds, measured both in relation to market indices and to their fellow funds. Such persistence proves to be significant, though it is predominantly a short-run phenomenon. Adopting the argot of the sports world, we label funds that deliver sustained short-run superior performance as having 'hot hands.'[3] Applying conventional Jensen and Sharpe measures, we measure performance of ex-ante investment strategies that exploit the identification of funds with hot hands. We find statistically significant potential for superior performance, with economically large risk-adjusted excess returns up to 10% per annum. We can also identify ex-ante underperformers with substantial negative excess returns.[4]

These results are striking since our methodological framework is simple and widely used in the literature, as in Grinblatt and Titman (1989). The results are not sensitive to the use of more-sophisticated benchmarks of performance proposed in the literature. Moreover, they hold up in the post-sample years of 1988 and 1989.

We evaluate the time horizon over which past performance is relevant. If the evaluation period is too short, the signal of superior performance due to skill is lost in the noise from chance factors. If the evaluation period is too long, the salience of hot hands diminishes. The strongest results appear when the evaluation period is one year, which is consistent with the lag-length beyond which the partial autocorrelations in excess returns are no longer significantly different from zero.

The paper is organized as follows. Section II outlines the hypotheses and presents the statistical test results that identify short-run persistence of performance, both superior and inferior. Section III demonstrates

---

[3]Camerer (1989) finds that point spreads for betting on basketball games are consistent with bettors believing in a 'hot hand' among professional basketball teams. His analysis of point spreads suggest that bettors respond too strongly to winning and losing streaks. However, bookmaker's commissions preclude profiting from Camerer's findings.

[4]Grinblatt and Titman (1987) also report performance persistence of the worst performers in their sample. However, the magnitudes of inferior performance for their constructions are much smaller than we find. Persistent inferior performance, of course, does not generate any exploitable investment strategy since open-end funds cannot be sold short, nor does it immediately reject the efficient markets hypothesis, since poor performers may be churning or otherwise building up expenses.

that hot hands in mutual funds can be exploited to achieve superior investment strategies. Section IV establishes the robustness of the findings. Section V concludes.

II. Properties of Mutual Fund Performance

A finding that recent performance can predict which funds are likely to perform well in the future would be inconsistent with the usual null hypothesis of efficient markets and martingale equity prices.[5] We analyze time-series characteristics of mutual-fund excess returns in this regard.

II.1 Hypotheses

We assess fund performance using the familiar market model applied to excess returns:

$$(R_{it} - R_{ft}) = \alpha_{it} + \beta_i(R_{mt} - R_{ft}) + \epsilon_{it}, \quad i=1,\ldots,N, \quad t=1,\ldots,T. \tag{1}$$

Here we have data over T time periods for N funds and where

R_{it} = the return by fund i over quarter t, net of all fees and assuming dividend reinvestment;

R_{ft} = the risk-free return over quarter t (which we proxy by the yield on 90-day U.S. treasury bills);

α_{it} = Jensen's alpha: measure of superiority of fund i in period t relative to the benchmark portfolio m in a mean-variance framework;

β_i = 'beta' of fund, which we assume to be time-invariant for convenience: measures systematic risk of fund i within the Capital Asset Pricing Model (CAPM);

R_{mt} = the return to the market (benchmark) portfolio over quarter t; and

ϵ_{it} = ex-post idiosyncratic component (error) of the return, which would be unpredictable under a joint hypothesis of the CAPM and the EMH.

---

[5]Though recent papers by Fama and French (1988) and Poterba and Summers (1988) suggest mean reversion in equity indices, the presumption of *unpredictable* excess returns on equity portfolios remains a useful starting point. Moreover, Richardson (1989) suggests that the recent evidence for mean reversion is flawed because the correlation among the serial correlation estimates and the jointness of returns across holding periods has been ignored.

We are interested in the dynamic properties, if any, of the α-parameter for mutual funds. The traditional null hypothesis is that α_{it} is ex-ante unpredictable by investors:

H1: $\alpha_{it|t-1} = 0$, for all i.

Here the notation '$|t-1$' indicates the expectation of the variable conditional on information available through time t-1. An alternate hypothesis would be that some funds have a constant nonzero ex-ante excess performance:

H2: $\alpha_{it|t-1} = \mu_i$, $\mu_i \neq 0$ for some i.

Typically, the alternate hypothesis of interest is superior performance by some funds, i.e., $\mu_i > 0$. Our evidence on H2 versus H1 is an update, with some refinements, of earlier tests.

We move beyond the earlier literature by conjecturing that the conditional mean is nonzero and time-varying:

H3: $\alpha_{it|t-1} = \mu_i + f_i(\alpha_{t-j};\ j>0)$, $f_i(\cdot) \neq 0$ for some i and some t.

Note that even if the unconditional mean, μ_i, is zero, we reject H1 as long as the conditional prediction, $f_i(\cdot)$, is nonzero for some t. H3 admits funds that have hot hands, that is, funds that are expected to be superior performers in the near term. For convenience in discussion, consider the special case of H3 when α follows a mean-zero univariate moving-average process of order J (MA[J]):

H3A: $\alpha_{it|t-1} = \sum_{j=1}^{J} \eta_{ij} u_{i,t-j}$.

Here the η's are moving-average weights, and u_{it} are the innovations driving α_{it}. While direct tests with a specific parametric model of hot hands like H3A can be powerful in discriminating against H1 or H2, the tests may be grossly misleading if the specialization of $f_i(\cdot)$ is inaccurate.

II.2 Data Sources and Sample Selection

Our benchmark ('market') portfolios are the equally-weighted portfolio of mutual funds in our sample (EWMF) and the portfolio indexed to the Standard & Poor's 500 (SP500). Treasury bill yields (our proxy for the risk-free/zero-beta return) and dividend-adjusted returns for SP500 are obtained from Ibbotson and Sinqeufield (1989). In our description of the results, we focus on the well-known SP500 benchmark though the results with EWMF are essentially the same.

We restrict our attention to no-load equity funds that have a growth objective (i.e., funds that identified themselves as seeking growth, aggressive growth, or growth and income). Our focus on equity funds makes reasonable our reliance on equity portfolio benchmarks. We concentrate on no-load funds because the transactions costs associated with investing in (and switching between) such funds are trivially different from zero, which is convenient for the switching strategies we consider. (We ignore tax consequences.)

Our returns data are net of management fees and assume that all dividends are reinvested. For the period 1974:4 to 1984:2, the returns were obtained from CDA Investment Technologies, Inc. of Silver Spring, Maryland. This data source is also used by Grinblatt and Titman (1989). Returns for the period after 1984:3 were taken from quarterly reports published in *Barron's* of data collected by Lipper Analytical Services, Inc. of New York City. We cross-checked the data for the 1982:2-84:2 period when the two sources overlap: the overlapping data agree with each other.

The sample of 96 funds is a subset of the 157 funds that met our growth strategy criterion in the *Barron's* listing in 1982. We dropped 61 funds: 10 funds adopted a load before to 1988, 21 were not in the CDA database at all, another 24 started in the CDA database only from 1984, and 6 were merged into other funds. The *Wiesenberger Investment Survey* suggests that our sample contains approximately 75% of the universe of no-load growth mutual funds during the 1970s; the proportion falls to about 50% by 1988.

In the post-1982 period, no fund disappears from the Lipper reports except because of merger; survivorship bias, if any, affects only the pre-1982 sample since the CDA database only includes funds that remain by 1984. Note, however, that the funds most likely to not have survived the entire sample period and therefore to be excluded from our sample are those that exhibited persistently poor performance. Such a sampling bias may lead to false rejections of H1 relative to H2 with the use of a benchmark like the SP500 (though it is less likely to do so with the EWMF which relies on the funds' sample itself). However, the same sampling bias will underestimate performance persistence and therefore tests unadjusted for selection bias will be biased in favor of H2 relative to H3. Since H3 is of main concern to us, our adjudications below in favor of H3, if affected at all, are probably understated.

II.3 Permanent Performance Persistence: H1 vs H2

In the appendix, we provide summary statistics and basic results that are directly comparable with the earlier literature on mutual funds. Briefly, the mutual fund betas are distributed around unity, and the majority of the estimated individual α's (excess returns) are not significantly different from zero. Joint tests for zero α are complicated by the fact that the number of funds, 96, exceeds the number of observations per fund, 54. We provide one battery of tests using a common factor model that relaxes the widely used but unconvincing assumption of cross-sectionally uncorrelated market-model errors (see Ippolito (1989, p.7 and table 1) or Grinblatt and Titman (1989)). In all the tests, the joint hypothesis of zero α's is easily rejected for our sample.

However, feasible investment strategies — not reported — that exploit the rejection of H1 in favor of H2 do *not* generate significant excess returns (either statistical or economic), a finding similar to Grinblatt and Titman (1987, table 9).[6] In sum, while we can statistically reject H1 in favor of H2, this appears to have little practical consequence.

II.4 Short-Run Performance Persistence (Hot Hands): H2 vs. H3

Let us turn to the statistical evidence on H2 versus H3, that is, between a constant α versus a time-varying α for individual funds. (Note that since H1 is a special case of H2, the rejection of H2 in favor of H3 also rejects H1). In our jargon, we assess the evidence for funds with hot hands. Of course, finding statistical evidence for hot hands need not imply that economically worthwhile investment strategies exist, but without the possibility for practical exploitation the phenomenon is far less interesting and less threatening to the EMH. We defer to section III the assessment of economic gains from identifying funds whose hands are hot.

II.4.1 Autocorrelation in market-model residuals. Consider the residuals, e_{it}, from estimating the market-model equation (1). Under H2, e_{it} is white noise. Under H3, the residuals will be autocorrelated. Under H3A, for in-

---

[6]Even allowing short sales of mutual funds, which is not possible in practice, does not lead to excess returns for strategies based on exploiting H2.

stance, e_{it} is the sum of an MA(J) process (our model for α) and white noise, which is also an MA(J) process.[7]

For any fund with T observations, an omnibus test of H2 can be based on the modified Q-statistic — see Harvey (1981, p.211):

$$Q = T(T+2)\sum_{j=1}^{L}[\hat{\rho}_j^2/(T-j)], \tag{2}$$

where $\hat{\rho}_j$ is the estimated residual autocorrelation at lag j. The Q-statistic tests the hypothesis that all of the autocorrelations of a series up to lag L are zero; its asymptotic distribution is χ^2_L. In our application, the Q-test may have low power because the autocorrelations will be close to zero, even if the η's under H3A are non-zero, if the variance of ϵ is much bigger than that of u.

We set L = 12, allowing for correlations up to three years. The Q-statistics for each fund appear in the last column of table 1. Of the 96 funds, 30 have Q-statistics significant at the 10% level.[8] A joint test of zero first-order autocorrelation in the residuals, $\hat{\rho}_{(i)1}$, of all funds is:

$$Q_1 = T\sum_{i=1}^{N}\hat{\rho}^2_{(i)1}. \tag{3}$$

Under H2, Q_1 has an asymptotic χ^2_N-distribution. We use this technique to construct Q_K statistics that include higher-order autocorrelations. Table 1 reports the results up to K = 8. The p-values of the Q_K statistics are close to zero: we can easily reject the null hypothesis of H2 in favor of H3.

For a practical exploitation of short-term persistence in performance, the approximate autoregressive order of Jensen's α under H3 is of interest. The order indicates the relevant time period for predicting future performance. Preliminary inferences can be based on examining the sums of squared *partial* correlations:

$$q_k = T\sum_{i=1}^{N}\hat{\rho}^2_{(i)kk}, \tag{4}$$

where $\hat{\rho}_{(i)kk}$ is the estimate of the k[th] partial autocorrelation in the residuals of fund i. Under the null hypothesis that an autoregression of order (k-1) or less fits the residuals, q_k is asymptotically distributed χ^2_N.

---

[7]The specific time-series process of e_{it} is sensitive to the exact specification appropriate under H3. Thus, for an autoregressive process of order J for α_{it}, instead of the MA(J) in H3A, e_{it} will follow an autoregressive moving-average process of order (J,J).

[8]The 10% significance level was chosen because this test has low power: we are trading off some type I risk for type II risk.

The pattern of q_k's in the lower panel of table 1 (which have low p-values up to $k=4$ and high p-values for q_5 through q_8) indicate that an AR(4) process adequately approximates the time-dependence in the market model residuals. Practically, performance information from the most recent four quarters appears sufficient for efficient prediction about future performance, which is consistent with the findings of ex-ante strategies detailed in section III. For now, we consider further statistical tests for hot hands.

II.4.2 Methods for Direct Assessment of Short-Run Persistence in α's. A direct assessment of performance predictability from period A to period B examines the relation between the two sets of $\hat{\alpha}$'s. This approach bypasses the problem due to a large variance of ϵ (which affects the previous tests based on residual autocorrelation) since it looks at deviations from the cross-sectional-mean $\hat{\alpha}$ (which may vary considerably over time). However, a direct cross-sectional regression between sets of $\hat{\alpha}$'s would have disturbances that are correlated across funds.[9] Thus, we employ the time-series regression approach discussed in Grinblatt and Titman (1989) as well as contingency table analysis.

The time-series approach recognizes that the slope-coefficient relating $\hat{\alpha}$'s in period B to those of period A is a weighted average of the $\hat{\alpha}_B$'s with weights that are proportional to the deviations of the $\hat{\alpha}_A$'s from their period-A mean. Thus, the slope-coefficient is equivalent to the Jensen measure of a self-financing portfolio of funds formed by choosing weights proportional to the deviations of the $\hat{\alpha}$'s from the mean of period A. We compute the excess returns to this portfolio during period B, say $\mathbf{R}_{pB}$. Let $\mathbf{R}_{mB}$ denote the corresponding vector of excess returns on the market portfolio during B. The t-statistic of the intercept from the time-series regression of $\mathbf{R}_{pB}$ on $\mathbf{R}_{mB}$ tests the hypothesis that α-performance in period A is correlated with α-performance in period B.

For a specified length of subperiod, T_s, our sample provides $M = T/(T_s+1)$ nonoverlapping pairs for study. A specific alternate hypothesis, say H3B, is that α-persistence is positive (hot hands) and the persistence relation is stable across subperiods. For this case, we perform the following regression:

---

[9]Another complication is that we do not know the true α, but only have a noisy estimate $\hat{\alpha}$ instead, which leads to errors-in-variables problems.

(5) $$R_{pt} = a + \sum_{i=1}^{M} b_i \delta_{it} R_{mt} + w_t, \quad t = 1,2,...,T,$$

where

R_{pt} = return in time t of the weighted-average portfolio whose weights depend on the subperiod to which the observation belongs,

a = intercept that measures α-persistence,

M = number of subperiods under consideration,

b_i = regression coefficient that is not of direct interest,

δ_{it} = dummy that is unity if t belongs to subperiod i and is zero otherwise, and

w_t = regression error that is assumed homoscedastic for all t.

In the case of H2 versus H3B, the t-statistic of the intercept, a, provides the test.

For contingency table analysis, consider $\hat{\alpha}$-quartiles for periods A and B. A 4x4 table is constructed such that the cell (j,k) contains the number of funds that fall into the j^{th} quartile of period A and the k^{th} quartile of period B. This method has several appealing features:

- Like the regression approaches, it focuses on the deviations of the performance measures from the cross-sectional mean at any point in time.
- It allows us to compare effects across performance levels. That is, we can look directly at whether persistence seems to be more evident among poorly performing funds than among funds that perform well, or vice versa.
- Nonlinearities in the relation may be uncovered. Also we avoid the risk of incorrect inference in the traditional framework because of possibly leptokurtic (fat-tailed relative to Gaussian distribution) returns. (This risk may be substantial; see Affleck-Graves and McDonald (1989).)

Of course, the contingency table analysis will have low power relative to a *correctly* specified parametric method.

We use the γ-statistic proposed by Goodman and Kruskal (1954) as a measure of ordinal association in the contingency table:

(6) $$\gamma = (P - Q) / (P + Q)$$

Here P is the number of concordant pairs of observations (that is, the number of paired observations where one member falls into a higher quartile in both periods); Q is the number of discordant pairs (i.e., one member falls into a higher quartile in period A and the other member is higher in period B). Observation pairs tied in one or both periods are ignored. Asymptotic tests use the asymptotic variance of γ given by Goodman and Kruskal (1972).

We analyze α-persistence for different subperiod lengths. The relation between the magnitude of the persistence and the length of the subsample will depend on the validity of H2 versus H3. If H2 is true, then the longer the period used to estimate α, the smaller the sampling variance of the estimate and thus the stronger the relation between the $\hat{\alpha}$'s from different periods. On the other hand, under H3, if the unconditional mean of α is zero but its conditional mean is time varying, then the relations will vary in a complicated manner but decay exponentially after some maximum subperiod length. We estimate α's over one-, two-, three-, and four-year periods, and over the half-sample (seven-year) periods. Since our subperiods do not overlap, there are twice as many subperiods for the one-year computations as for the two-year, and so on.

II.4.3 Results on Direct Assessment of Short-Run Persistence in α's. The regression results and γ-statistics that employ the methods of section II.4.2 are presented in table 2. The results of $\hat{\beta}$-persistence are shown for comparison. The tests show that $\hat{\alpha}$'s and the $\hat{\beta}$'s display similar significant persistence for the one-year periods: the γ-statistic for $\hat{\alpha}$'s is 0.32 and that for $\hat{\beta}$'s is 0.40. The $\hat{\beta}$-persistence is expected. That of the $\hat{\alpha}$'s is surprising, given that the EMH suggests zero persistence. However the magnitude of $\hat{\alpha}$-persistence, when assessed by γ- or t-statistics, diminishes to zero by about three- year estimation periods. In contrast, the persistence in $\hat{\beta}$ is long-lived, as expected.

In table 2, the t-statistics and the γ's reject H2 in favor of H3, which is consistent with the earlier analysis of residual autocorrelations. The decay of $\hat{\alpha}$-persistence with increasing estimation-interval length suggests that the unconditional mean of α's is zero, since otherwise some persistence, possibly attenuated, would be observed for long periods.

We can also examine the relation between $\hat{\alpha}$'s computed over one-year periods with different intervals between the one-year estimates. Such an examination provides inferences about the efficient evaluation-period length for forecasting performance. (The earlier tests based on partial autocorrelations in table 1 are analogous.) Under H2, the relation between the $\hat{\alpha}$'s should not be affected by the interval between the years of the estimates; under H3, the relation should be sensitive to the interval. The procedures used for table 2 were applied to study annual α-persistence between: year t versus t+2, year t versus t+3, year t versus t+4, and year t versus t+5. The results appear in table 3. (Year t versus t+1 is repeated from table 2 for convenient comparision.)

The results in table 3 support H3 (short-run persistence) relative to H2 (long-run persistence) since the $\hat{\alpha}$-persistence, both as measured by the regression intercept and by γ based on the contingency tables, diminishes substantially when we allow a nonzero interval between the estimation years.

In summary, the evidence in tables 1, 2, and 3 indicates statistically significant persistence between performance from one period to the next. The persistence is greatest for an assessment period of one year, from which we seem to obtain sufficient information on performance to overcome noise yet retain enough recency to be relevant. The persistence fades away in the long run, which is consistent with a hot-hands phenomenon.

III. Performance of Strategies Based on Hot Hands

To learn whether hot hands are economically important, we generate comingled portfolios from open-end mutual funds using historical information on superior short-term performance. Our evaluation is based on Jensen's α (see equation (1)) and on the difference in Sharpe's measure (excess return per unit of standard deviation risk) between the hot hands and benchmark portfolios. The significance of the difference in Sharpe's measure is assessed by normalizing it by its asymptotic standard deviation (computed following Jobson and Korkie (1981)); the ratio, denoted as z-statistic, is asymptotically distributed standard normal.

As is well known, Jensen's α or differential Sharpe's measure will provide ambiguous results if our benchmark portfolio does not lie on the efficient frontier — see Roll (1977, 1978). Simple equity-market indices, which were common benchmarks in the 1970s, do not seem to lie on the mean-variance efficient frontier — see Grinblatt and Titman (1987), De Bondt and Thaler (1989), and references therein. Substantial diversification gains can accrue from including real estate in portfolios — see Firstenberg, Ross, and Zisler (1988) for recent evidence. Likewise, a case can be made for including foreign securities. Clearly, our benchmarks are unlikely to be globally efficient. Reassuringly, our results withstand the scrutiny in section IV of simulations and alternative benchmarks. Finally, since we reuse the sample of section II rather than test on a new sample, our results on hot-hand strategy gains should be interpreted cautiously. (Though, again, we are reassured by some out-of-sample evidence presented in section IV.)

Practically, following Dybvig and Ross (1985), we assess whether a hot-hands strategy improves performance unambiguously *relative* to the benchmark in a mean-variance framework (given the choice of the risk-free asset). While the success of the hot-hands strategies need not imply a rejection of the EMH, economically large and significant gains should raise questions about the prior mutual-fund studies' apparent support for the EMH.

Consider a strategy that invests in a equally-weighted mix of mutual funds that is updated at the end of each holding period. To exploit hot hands, the mix of mutual funds for a holding period is based on the top performers in the most recent evaluation period. The notation mEnH for a strategy indicates a m-quarter evaluation period and a n-quarter holding period: for example, 4E8H will indicate a procedure based on the most-recent 4-quarter performance, with selections updated every 8 quarters. The simple net return (a naive performance measure justified by a prior of unit beta for each fund) from the evaluation period is the criterion for fund inclusion. In results not reported, we find that selection based on an estimate of Jensen's α from the evaluation period performs similarly or slightly better; selection based on Sharpe's measure, however, generates performance that is insignificantly different from the benchmarks.

In the tables that follow, we report the net quarterly return, the Sharpe's measure (asterisks indicating values that are statistically different from those with the EWMF portfolio), and Jensen's alpha with its t-statistic to assess the magnitude of excess returns.

In table 4, top panel, we focus on a best-fund strategy in which the evaluation period and the holding period are equal, mEmH, ranging from one quarter to 12 quarters. An annual horizon, 4E4H, appears best, which is consonant with the earlier statistical results in tables 1 and 3. The α-estimate for 4E4H is statistically significant and indicates an excess annualized return greater than 10%. The Sharpe's measure for 4E4H is also statistically significantly different from that of the benchmarks. The middle panel in table 4 reports on the best-fund strategy performance with varying holding periods and a 4-quarter evaluation period. The bottom panel in table 4 reports on the performance variation obtained when the number of top-performing funds that are selected varies. Including more funds lowers the extent of superior performance, by both the Jensen's and Sharpe's measures. (However, the statistical significance of α remains invariant when assessed with the EWMF benchmark, perhaps because idiosyncratic variation decreases.) Any diversification gains from increasing the number of included funds appear to be offset by reduced selectivity, at least with naive equal weighting.

Overall, the best-fund 4E4H strategy based on hot hands leads to a statistically significant Jensen's α, with annualized values of 10%, a remarkable record of excess performance. A best-α fund strategy finds similar excess returns of 5-12% per year. In figure 1, we show the average relation between percentile ranks of one year with the next. (Since the average relation is shown, the percentiles for period t+1 do not span 0 to 100.) A best-fitting regression line is shown in the figure for reference. We observe that performance persistence is not restricted to the extreme ranks but is uniformly distributed. Thus, a fund in the second-best decile is more likely to outperform, in the next year, a fund in the third-best decile, and so on.

Of course, a hot-hands selection strategy does not guarantee superior performance every year. Figure 2 shows the distribution of the year-by-year decile ranks of the best-fund 4E4H strategy. While the top decile is the mode (with 5 outcomes in 12 years), we do observe a number of below-median rankings.

We can examine the selectivity of the hot hands approach in distinguishing prospective winners from losers by comparing the best fund strategy with a worst fund strategy. Table 5 reports the results for the strategy that selects the worst performer annually using a annual evaluation period (4E4H). (The results for the best fund strategy differ slightly from table 4 because of different sample periods.) The worst-performer fund consistently generates sigificantly negative excess returns that are below -9% per annum. The counterpart to hot hands, what we might call icy hands, appears to be an equally strong effect. A hypothetical short position in the worst-performer fund combined with a long position in the best-performer fund generates large Jensen's α's and differential Sharpe's measures, which are statistically significant for other mEnH combinations too (not reported). For the 4E4H combination, we observe risk-adjusted annualized excess returns of +21% for this long-and-short strategy. Since the market-model residual variances for the hot-hands strategy vary across holding periods, selected heteroscedasticity-corrected estimates are also shown. The overall results and inferences stay unaffected by corrections for heteroscedasticity.

Table 5 also shows selected results for the subperiods 1976-81 and 1982-87. The Jensen's α is much higher for the best-performer fund during 1976-81 than for 1982-87. In contrast, the worst-performer fund does much worse during 1982-87. Thus, the selectivity of the hot-hands strategy is high in each subperiod: for instance, Jensen's $\hat{\alpha}$ (based on the SP500) for the best-worst strategy has a p-value of less than 5% in each period.

IV. Robustness of Hot-Hands Finding

We consider three explorations on the robustness of the findings.[10] First, using simulations, we shed light on whether the excess returns that we find are attributable to (a) simply selecting good funds rather than selecting the right fund at the right time, or (b) a spurious interaction between our selection strategy and the structure of equity returns in the 1974-87 sample period. Second, we consider a multiple portfolio benchmark that accounts for documented anomalies when using the traditional indices. Third, we update the best-fund

---

[10]In results not reported, we find that our findings are similar between bull and bear markets.

Also, we evaluated a specific implementation of H3 where Jensen's α follows an autoregressive model:

(*) $\alpha_{it} = a + \phi\alpha_{i,t-1} + n_{it}$.

Equation (1), the market model, was estimated by applying the Kalman filter. We specified a autoregressive representation for β as well:

(**) $\beta_{it} = b + \theta\beta_{i,t-1} + v_{it}$.

For convenience, we focused on autogressive models of order one. (Second-order specifications gave similar results). We imposed a tight prior on β centered at unity, and a loose prior on α centered at zero. The variance of ϵ, the error term in equation (1), was set at 80% of the estimated residual variance from ordinary least squares.

For a sample of 10 funds out of 96, we explored a grid of choices for the autoregressive parameters (i.e., for ϕ_1 and θ_1 in equations (*) and (**) respectively) and for the variances of n and v. Typically, $\sigma_n > \sigma_v$ while $\phi_1 > \theta_1$. A satisfactory specification was then used for all 96 funds. For each fund, at each quarter from 1976-87, we computed the one-step prediction of α, $\alpha_{it|t-1}$. The strategy was to select the funds with the highest $\alpha_{it|t-1}$ for each period t. Our time-varying parameterization did not lead to superior performance. Within the time-varying coefficients model, we also analyzed a selection strategy that allowed for estimation risk without success.

strategy's results with post-sample data from 1988-89.

IV.1 Simulations: Selectivity vs. Timing and Sample Artifacts

Our tests of permanent performance persistence (H2) versus short-run persistence (H3) favored the latter. This distinction between pure selectivity and timing selectivity is a central issue in this paper. We report on simulations that provide further evidence on this issue.

Consider the two-fund 4E4H strategy, which selects, each year, the two funds that performed the best in the most recent year (based on Jensen's α). In our sample, 4E4H picks 19 different funds during 1976-87. (The maximum possible number of funds that could be picked is 24.) We simulate 5000 times the performance of a portfolio that includes, each year, two funds randomly chosen from among the 19. The probability of including a particular fund in any year is set equal to the relative frequency observed in our actual sample.[11] The percentiles from the simulations of Jensen's α, Sharpe's measure, and average excess return are shown in table 6. The original observed value is indicated in each of the panels. Less than 5% of the simulations have values larger than the observed values from our 4E4H strategy. In fact, the central 80% of the distribution for each measure always includes the corresponding values of the benchmark portfolios. Thus, it appears unlikely that pure selectivity of funds is the source of the potential for risk-adjusted superior performance. We conclude that the hot-hands strategy displays timing ability (i.e., the ability to pick funds at a good time), quite apart from picking good funds.

Next, we evaluate the likelihood that the hot-hands findings were generated spuriously, because of a chance interaction of our selection procedure and of the time-series properties of equity returns during 1974-87. We generate 100 artificial portfolios, each of which is an equally weighted portfolio of 100 equities drawn randomly from the NYSE/AMEX (New York Stock Exchange and American Exchange) stocks on the widely used monthly returns tapes constructed by the Center for Research in Security Prices, University of Chicago. We apply the best-fund 4E4H strategy to this set of unmanaged portfolios/funds over the sample period identical to

[11]This approach amounts to sampling with replacement from the admissible funds. Simulations based on sampling without replacement give virtually identical results and are not reported.

that used in our analysis in section III. One hundred such simulations are carried out. With the unmanaged portfolios, we find that the 4E4H best-fund strategy does not generate excess returns; in fact, its average excess return (measured by Jensen's α) is slightly negative with the benchmarks of the P8 portfolios or the EWMF. With the SP500 benchmark, the 4E4H strategy's α is positive and significant but this is inconsequential because the *average* unmanaged portfolio also obtains a similar α — essentially, the SP500 is an inappropriate benchmark since the unmanaged portfolios are, on average, composed of equities from much smaller firms than those in the SP500 and thus the well-known size bias surfaces. We conclude that the hot-hand finding is extremely unlikely to be an artifact of the sample period returns.

IV.2 Alternate Benchmarks

In recent years, anomalies in the risk-return relations with common indices have been linked to firm size, dividend yield, and returns reversions. That is, the returns of small firm portfolios, portfolios of firms that pay high-dividends, or portfolios of firms that have performed very poorly recently, exhibit significant positive Jensen α's when common indices are used as benchmarks. Conceivably, therefore, our hot-hands results with the SP500 or the EWMF merely mimic such well-known phenomena.

We consider an eight-portfolio benchmark (denoted P8) that accounts for size, dividend, and mean-reversion anomalies, as described in Grinblatt and Titman (1987, 1989). Grinblatt and Titman argue convincingly that the P8 benchmark is preferable to other candidates, such as the equally weighted CRSP portfolio, the value-weighted CRSP index, or the factor portfolios discussed by Lehmann and Modest (1988). The performance of the 4E4H strategy relative to P8 is presented in table 7. Since the P8 returns were only available for the 1975-84 subperiod, we show comparable performance results for the subperiod with SP500. Results with the P8 benchmark leave the conclusions based on SP500 unchanged: we continue to obtain a significantly positive Jensen's alpha.[12] The substantial selectivity potential identified in the previous section does not appear to

---

[12] In results not reported, we found that the superior performance of the 4E4H strategy becomes insignificant when we use the benchmark of the ten-factor portfolios, F10, of Lehmann and Modest (1988). However, this deviant finding is readily explained by the observation of Grinblatt and Titman (1989, p. 396) that "In particular, funds that invest in large firms (which includes most funds) tend to exhibit negative performance with the EW and F10 benchmarks." The Grinblatt and Titman conclusion is based on finding that the F10 portfolios display size, dividend-yield, and beta-related pricing errors.

reflect the choice of an inappropriate benchmark.

IV.3 Best-Fund 4E4H Strategy in 1988-89

The results in sections II and III are based on the same dataset; hence their marginal reinforcement for the hot hands finding is less than additive. In future research, we plan to extend the examination to different mutual fund classes as well as to the unit trusts of U.K. The out-of-sample average quarterly returns for the best-fund 4E4H strategy, EWMF (average mutual fund), and SP500 are as follows:

- For 1988, 3.1%, 1.6%, and 2.4% respectively.
- For 1989, 5.9%, 4.0%, and 5.1% respectively.
- For 1988-89, 4.48%, 2.79%, and 3.77% respectively.

In the post-sample period, the 4E4H strategy continues to outperform the EWMF as well as the SP500. The hot-hands persistence holds up in recent years.

V. Concluding Discussion

We found a hot-hands phenomenon in net returns of no-load mutual growth funds during 1974-87. Specifically, mutual funds that perform well in the most recent year continue to be superior net performers in the near term (one to eight quarters). A best-fund strategy with an annual holding-period and an annual evaluation-period that exploits hot hands generates significant risk-adjusted excess returns of 10% per year. Icy hands, the negative counterpart to hot hands, also show up in our sample: funds that perform poorly in the most recent year continue to be inferior performers in the near term. This phenomenon is possibly even more significant in the statistical sense, though not exploitable. Since our hot-hands strategies are based only on knowledge of historical returns, the findings cast doubt on the weak-form efficient markets hypothesis and differ sharply from the established literature on mutual funds, which reports no exploitable opportunity to achieve risk-adjusted superior performance.

For the 1974-87 sample period, the significant performance of the hot (icy) hands strategy can not be explained either as a result of simply selecting superior (inferior) funds or as an artifact of the equity returns structure. There is a significant timing component. The superior performance of hot-hand strategies persists in 1988-89 and with alternative benchmarks, including an eight-portfolio benchmark that accounts for firm-size effects, dividend yields, and reversion in returns.

We leave unexplained the causes that underlie the observed time-decay in the capacity of superior performing funds. Plausible conjectures include:

- bidding away of superior analysts once they build a track record,
- excessive new funds flow to successful performers, leading to a bloated organization and fewer good investment ideas per managed dollar,
- loss of urgency and drive once reputation is established,
- market feel that is limited to specific circumstances, and
- rise in fees and salaries to capitalize on demands arising from recent successes.

More generally, there may be a life cycle for effective organizations: witness the decline of General Motors or Great Britain.

Like baseball teams and pop singers, stellar mutual funds typically fade away after a few years. Substantial gains are available from investing in the mutual fund equivalents of last year's pennant winners.

Appendix

Testing for Permanent Superior Performance: H1 vs. H2

Table A1, panel A, reports simple summary statistics for 90-day Treasury bills. The equally-weighted portfolio of mutual funds in our sample (EWMF) and the SP500 perform about equally.

For a test of H1 versus H2 (i.e., of some funds performing persistently differently from the benchmark), we estimate equation (1) (the market model in excess returns) with ordinary least squares using the entire sample period.[13] Table A1, panel B, reports the estimates and the t-statistics for each of the 96 funds in our sample. The estimated slope-coefficients, $\hat{\beta}$'s, are scattered about unity; the market model R-squared values (not reported) are around 0.8. The majority of the estimated intercepts, $\hat{\alpha}$'s, are not significantly different from zero.

The $\hat{\beta}$ for EWMF is 1.069, which is statistically not different from unity. The $\hat{\alpha}$ for EWMF is 0.123 with a p-value of 0.7:[14] hence H1 (i.e., the hypothesis of a zero α) is not rejected. This, however, is not the same as a test of all the α_i's being jointly equal to zero. A joint test can be constructed using the 'seemingly unrelated regression' framework, SUR. The null hypothesis of H1 asserts that $R\pi = 0$, where R is a restriction matrix that selects the α's and where $\pi = [\alpha_1, \beta_1, \alpha_2, \beta_2, ..., \alpha_N, \beta_N]'$ is the vector of stacked regression coefficients from each market model regression. It can be shown that

(A1) $g = (R\hat{\pi})'(RCR')^{-1}(R\hat{\pi})$

[13] An assumption of a constant fund beta during the sample period is unreasonable if the fund changes its strategy. Also, if the fund's strategy is to "time the market," then Jensen's α is not a consistent estimator of superior performance — see Henriksson (1984), although he uncovers little evidence of superior timing ability in his sample of 115 mutual funds.

[14] A p-value gives the probability of observing the estimated value under the null hypothesis. The p-value stays above 0.1 even if we correct for the substantial autocorrelation in the residuals indicated by the large Q-statistic.

is asymptotically distributed χ^2_{96} since we have 96 funds and one α-restriction per fund.[15] Here $C = \hat{\Sigma} \otimes (X'X)^{-1}$, $\hat{\Sigma}$ is an estimated variance-covariance matrix of market-model errors across funds, and X is the regressor matrix (a constant and the excess benchmark returns) which is identical across cross-sections. Unfortunately the g-statistic cannot be directly computed since we can't obtain $\hat{\Sigma}$ when the number of funds, N (=96), exceeds the number of time periods, T (=54). This problem is frequently encountered in the literature evaluating mutual funds. We consider three different methods, with increasing reasonableness in our judgement.

SUR Approach 1. Typically, — see Grinblatt and Titman (1989) and Ippolito (1989, p. 7) — zero covariances between market-model errors are assumed, which is a heroic assumption at best since the average sample cross-correlation between errors is 0.23. Under this assumption of cross-sectionally uncorrelated market-model errors, we reject the null hypothesis of a zero α since, among 96 funds in table A1, we observe 19 non-zero $\hat{\alpha}$'s at a p-value below 5%. Under H1, this outcome has a probability of less than 1%. This finding is similar to Ippolito (1989, table 1). More directly, the g-statistic with a diagonal $\hat{\Sigma}$ is 219, which has a p-value below 0.1% under H1 and thus favors H2.

SUR Approach 2. Another typical approach considers subsets of funds so that N < T. Under this approach, we randomly selected 40 funds and computed the g-statistic. Table A2, top panel, reports on the results from 100 repetitions. In each case, the null joint hypothesis of a zero α for every included fund can be rejected — similar rejection of H1 is obtained by Grinblatt and Titman (1989) in their subsample results for growth funds.

SUR Approach 3. Our preferred approach estimates a $\hat{\Sigma}$ matrix under less restrictive assumptions than Approach 1. As a first cut, we modified $\hat{\Sigma}$ to reflect the sample cross-correlation average of 0.23 between funds. The g-statistic, reported in table A2, lower panel, is computed to be 281, which has a near-zero p-

[15]For our situation, where we have identical constraints and identical regressor matrices for each fund, Laitinen's (1978) results indicate that reliance on the asymptotic χ^2 distribution, at any nominal significance level, may tend to reject a correct null hypothesis more frequently than it should. In our study this concern is mitigated because our test-statistic values are very large and imply nominal p-value levels below 0.1%.

value.[16] A more general method is to assume that the market-model-error correlations can be modeled as arising from an underlying common-factor model. The equation for the common factor model that is applied to market model errors is:

(A2) $\epsilon_{it} = \sum_{k=1}^{K} \phi_{ik} h_{kt} + v_{it}$,

where

ϵ_{it} = the market-model error for fund i during time t,

ϕ_{ik} = the loading on factor k for predicting error i,

h_{kt} = the value of factor k during time t,

v_{it} = the unique factor for fund i (i.e., is uncorrelated with the unique factors corresponding to the other funds), and

K = the number of common factors.

In matrix terms, we can write:

$$E = H\Phi + V.$$

H is the matrix of factor scores, and Φ' is the factor pattern. The factors are normalized to have unit variance and rotated to be uncorrelated with each other.

We estimate Φ and V assuming five factors (K=5) by iterated unweighted least squares.[17] Under the assumptions of factor analysis, $\Sigma = \Phi\Phi' + V^2$. Hence we obtain the desired $\hat{\Sigma}$. The g-statistic with this $\hat{\Sigma}$ is computed to be 496, which has a negligible asymptotic p-value under H1.[18]

---

[16]The g-statistics for assumed cross-correlations of +0.5, -0.25, and -0.5 appear also in table A2, lower panel. The results indicate that only substantial negative correlations between funds could fail to reject H1 in our sample; such a correlation pattern is contra-indicated both by the sample evidence and by common intuition.

[17]Maximum likelihood would require N<T.

[18]For widespread use of this approach, future research will have to establish small sample properties. Such assessment is probably best conducted by bootstrap simulations since sensitivity to significant leptokurtosis (which is widely recognized in equity returns) or skewness (which may arise from dynamic portfolio management with option-like position-taking) are of concern.

However, positive investment strategies -- not reported -- that exploit the rejection of H1 in favor of H2 do *not* generate significant excess returns (either statistical or economic), a finding similar to Grinblatt and Titman (1987, table 9).[19] The result -- failure to find significant ex-ante performance strategies despite statistically significant rejection of H1 -- is consistent with survivorship bias in the sample. In sum, while we can statistically reject H1 in favor of H2, that finding appears to have little practical consequence.

---

[19]Even allowing short sales of mutual funds, which is not possible in practice, does not lead to excess returns for strategies based on exploiting H2.

References

Affleck-Graves, J., and B. McDonald, 1989, " Nonnormalities and Tests of Asset Pricing Theories," *Journal of Finance*, XLIV, n. 4, 889-908.

Brealey, R. A., 1983, " Can Professional Investors Beat the Market," chapter 3 in *An Introduction to Risk and Return for Common Stocks*, 2nd edition, M.I.T. Press (Cambridge, Massachusetts).

Camerer, Colin, 1989, " Does the Basketball Market Believe in the 'Hot Hand'?" , *American Economic Review*, 79, n. 5, 1257-61.

Consumer Guide, 1988, *Top Performing Mutual Funds*, 483.

De Bondt, W. F., and R. H. Thaler, 1989, " Anomalies: A Mean-Reverting Walk Down Wall Street," *Journal of Economic Perspectives*, 3 (1), 189-202.

Dybvig, P. H., and S. A. Ross, 1985, " The Analytics of Performance Measurement Using a Security Market Line," *Journal of Finance*, 40, 401-416.

Fama, E. F., and K. French, 1988, " Permanent and Temporary Components of Stock Prices," *Journal of Political Economy*, 96, 246-273.

Firstenberg, P., S. Ross, and R. Zisler, 1988, " Real Estate: The whole story," *The Journal of Portfolio Management*, Spring, 22-34.

Goodman, L.A., and W.H. Kruskal, 1954, " Measures of Association for Cross-Classifications," *Journal of the American Statistical Association*, 49, 732-764.

Goodman, L.A., and W.H. Kruskal, 1972, " Measures of Association for Cross-Classifications, IV: Simplification of Asymptotic Variances," *Journal of the American Statistical Association*, 67, 415-421.

Grinblatt, M., and S. Titman, 1987, " A Comparison of Measures of Abnormal Performance on a Sample of Mutual Fund Returns," UCLA Working Paper #13-86.

Grinblatt, M., and S. Titman, 1989, " Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings," *Journal of Business* 62, 393-416.

Harvey, A.C., 1981, *The Econometric Analysis of Time Series* (Philip Allan Publishers, Oxford).

Henriksson, R., 1984, " Market Timing and Mutual Fund Performance: An Empirical Investigation," *Journal of Business*, 57, 73-96.

Ippolito, R. A., 1989, " Efficiency with Costly Information: A Study of Mutual Fund Performance, 1965-84," *Quarterly Journal of Economics*, 104, 1-23.

Jensen, M.C., 1968, " The Performance of Mutual Funds in the Period 1945-1964," *Journal of Finance*, 23, 389-416.

Jobson, J.D. and B. M. Korkie, 1981, " Performance Hypothesis Testing with the Sharpe and Treynor Measures," *Journal of Finance*, 36, 889-908.

Henriksson, R., 1988, " Market Timing and Mutual Fund Performance: An Empirical Investigation," *Journal of Business*, 57, 73-96.

Laitinen, K., 1978, " Why Is Demand Heterogeneity So Often Rejected," *Economic Letters*, 1, 187-191.

Lehmann, B. N., and D. M. Modest, 1988, " The Empirical Foundations of the Arbitrage Pricing Theory," *Journal of Financial Economics*, 21, 213-54.

Patel, J., R. Zeckhauser, and D. Hendricks, 1990, " Investment Flows and Performance: Evidence from Mutual Funds, Cross-Border Investments, and New Issues," forthcoming in a volume of the Center of Japan-U.S. Business and Economic Studies, New York University.

Poterba, J., and L. Summers, 1988, " Mean Reversions in Stock Prices: Evidence and Implications," *Journal of Financial Economics*, 22, 27-59.

Richardson, M., 1989, " Temporary Components of Stock Prices: A Skeptic's View," manuscript, Wharton School of Business, University of Pennsylvania.

Roll, R., 1977, " A Critique of the Asset Pricing Theory's Tests - Part I: On Past and Potential Testability of the Theory," *Journal of Financial Economics*, 4, 129-176.

Roll, R., 1978, " Ambiguity When Performance Is Measured by the Security Market Line," *Journal of Finance*, 33, 1051-1069.

Rugg, D., 1986, *Dow Jones-Irwin Guide to Mutual Funds* (3rd edition).

Sharpe, W., 1966, " Mutual Fund Performance," *Journal of Business*, 39, 119-138.

Shawky, H. A., 1982, " An Update on Mutual Funds: Better Grades," *Journal of Portfolio Management*, Winter, 29-34.

Treynor, J., 1965, " How to Rate Management of Investment Funds," *Harvard Business Review*, 44, 131-136.

Tversky, Amos, and Daniel Kahneman, 1971, " Belief in the Law of Small Numbers," *Pyschological Bulletin*, 76, August, 105-110.

Wiesenberger, A., 1974-1987, *Investment Company Survey* (A. Wiesenberger & Co., New York).

</ref_section>

Table 1

Persistence in Market Model Residuals

Quarterly Returns: 1974Q4 - 1988Q1

Joint test of zero autocorrelations

(Assessing persistence based on equation (5))

	Autocorrelations Up To Lag K							
	K = 1	K = 2	K = 3	K = 4	K = 5	K = 6	K = 7	K = 8
Q_k-statistic[a]	119	229	396	538	606	703	793	881
p-value	5%	3%	0%	0%	0%	0%	0%	0%

Joint test of zero partial autocorrelations

(Inference on approximate order of autoregression based on equation (6))

	Partial Autocorrelation at Lag k							
	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8
q_k-statistic[b]	119	90	153	120	74	71	77	66
p-value	5%	65%	0%	2%	95%	97%	91%	99%

[a]The Q_k-statistic is $[T \sum_{i=1}^{N} \sum_{k=1}^{K} \hat{\rho}^2_{(i)k}]$, where $\hat{\rho}_{(i)k}$ is the autocorrelation at lag k in the market model residuals of fund i. It is asymptotically distributed χ^2_{96K} if the true autocorrelations are zero.

[b]The q_k-statistic is $[T \sum_{i=1}^{N} \hat{\rho}^2_{(i)kk}]$, where $\hat{\rho}_{(i)kk}$ is the partial autocorrelation at lag k in the market model residuals of fund i. It is asymptotically distributed χ^2_{96} if the autoregressive order is less than k.

Note: In the lower panel, the small q_k values beyond lag 4 suggest that the market model residuals, and hence Jensen's α, may be approximated by a fourth-order autoregression.

Table 2
Persistence Measures Between Estimated Alphas
and Betas for Different Sub-Period Lengths

Persistence in Alpha

Length of Estimation Sub-Periods	Regression Measure (t-statistic)	Gamma Measure [probability]
One Year	0.30 (3.94)**	0.32 [0.00]
Two Years	0.25 (3.29)**	0.24 [0.00]
Three Years	0.22 (2.11)*	0.17 [0.06]
Four Years	-0.00 (-0.01)	0.10 [0.23]
Sample Halves	-0.13 (-1.16)	-0.03 [0.57]

Persistence in Beta

Length of Estimation Sub-Periods	Regression Measure (t-statistic)	Gamma Measure [probability]
One Year	0.40 (5.89)**	0.40 [0.00]
Two Years	0.54 (5.66)**	0.45 [0.00]
Three Years	0.56 (6.64)**	0.54 [0.00]
Four Years	0.52 (5.46)**	0.50 [0.00]
Sample Halves	0.39 (4.86)**	0.45 [0.00]

* = p-value is below 0.05. ** = p-value is below 0.01.

Notes: Time-series regression t-statistics are in parentheses.
The p-values for the gamma measures are in parantheses.
The regression coefficients and their associated t-statistics are estimated using the time-series technique described in section II.4.2; the gamma statistic is also discussed in the same section.

Table 3

Persistence Measures for Alphas and Betas for Varying Intervals, One-Year Estimation Periods

Persistence in Alpha

Interval	Regression Measure (t-statistic)	Gamma Measure [probability]
Zero Years	0.30 (3.94)**	0.32 [0.00]
One Year	0.18 (3.17)**	0.11 [0.03]
Two Years	0.06 (0.82)	0.11 [0.03]
Three Years	-0.04 (-0.49)	-0.02 [0.64]
Four Years	-0.07 (-1.23)	-0.13 [0.97]

Persistence in Beta

Interval	Regression Measure (t-statistic)	Gamma Measure [probability]
Zero Years	0.40 (5.89)**	0.40 [0.00]
One Year	0.28 (4.19)**	0.28 [0.00]
Two Years	0.16 (2.48)*	0.35 [0.00]
Three Years	0.23 (4.80)**	0.32 [0.00]
Four Years	0.26 (5.66)**	0.28 [0.00]

* = p-value is below 0.05. ** = p-value is below 0.01.

Notes: Time-series regression t-statistics are in parentheses.
The p-values for the gamma measures are in brackets.
The regression coefficients and their associated t-statistics are estimated using the time-series technique described in section II.4.2; the gamma statistic is also discussed in the same section.

Table 4

Performance Results of Portfolios Formed on the Basis of Past Performance
Quarterly Sample Period: 1978Q1 - 1987Q4

			Jensen's Alpha (%) Benchmarks			
			S&P 500		EWMF[a]	
Strategy[b]	Mean Return	Sharpe's Measure[c]	Value	t-statistic	Value	t-statistic
Benchmarks						
S&P 500	1.76	0.20	—	—	—	—
EWMF[a]	1.73	0.18	—	—	—	—
Variable Holding and Evaluation Periods, Top Fund Included						
1E 1H	2.93	0.19	0.55	0.35	0.70	0.49
2E 2H	2.28	0.16	0.09	0.06	0.28	0.21
4E 4H	4.38	0.37*	2.60	2.16*	2.64	2.63*
8E 8H	2.83	0.19	0.45	0.34	0.54	0.53
12E 12H	3.75	0.26	1.34	1.09	1.44	1.54
Variable Holding Period, Annual Evaluation Period, Top Fund Included						
4E 1H	2.89	0.25	0.75	1.01	0.89	2.00
4E 2H	1.64	0.13	-0.17	-0.12	-0.05	-0.04
4E 3H	2.96	0.22	0.82	0.66	0.91	0.89
4E 4H	4.38	0.37*	2.60	2.16*	2.64	2.63*
Annual Holding Period, Annual Evaluation Period, Number of Funds Included Varies						
4E 4H 1 Fund	4.38	0.37*	2.60	2.16*	2.64	2.63*
4E 4H 2 Funds	3.87	0.33*	1.98	1.87	2.05	2.45*
4E 4H 5 Funds	3.55	0.30*	1.52	1.68	1.61	2.57*
4E 4H 10 Funds	3.24	0.29**	1.20	1.62	1.32	2.86**

* p-value is below 5%; ** p-value is below 1%.

[a]EWMF is the equally-weighted portfolio of all mutual funds in our sample.

[b]The notation mE nH indicates an m-quarter evaluation period and an n-quarter holding period. For example, 4E 2H indicates a portfolio for which the fund selections are based on an evaluation interval of the most recent 4 quarters and the selections are updated every 2 quarters.

[c]Sharpe's measure is the mean of the quarterly returns divided by the standard deviation of the quarterly returns. For each of the Sharpe's measures, a z-statistic was calculated to test the significance of the difference in Sharpe's measure between the strategy and the EWMF. Significance of the z-statistic is indicated by asterisks.

Table 5

Performance of Portfolios of Worst Fund Versus Portoflios of Best Fund

Annual Holding Period, Annual Evaluation Period, Best or Worst Fund

			Jensen's Alpha (%) Benchmarks			
			S&P 500		EWMF[a]	
4E 4H Strategy[b]	Mean Return[c]	Sharpe's Measure[d]	Value	t-statistic	Value	t-statistic
Sample Period: 1976Q1 - 1987Q4						
Worst Fund	-0.55	-0.05**	-2.26	-2.15*	-2.51	-2.76**
Worst (corrected)[e]	-0.55	-0.05**	-0.76	-1.12	-1.76	-2.19*
Best Fund	4.78	0.40*	3.10	2.67*	2.77	2.88**
Best (corrected)[e]	4.78	0.40*	2.56	2.56*	2.63	2.99**
Best - Worst	5.33	0.52*	5.36	3.55**	5.28	3.49**
Sample Period: 1976Q1 - 1981Q4						
Worst Fund	1.49	0.21	0.99	1.19	-0.12	-0.15
Best Fund	6.38	0.44*	5.35	3.40**	2.82	2.38*
Best - Worst	4.88	0.47	4.35	2.62*	2.94	1.91
Sample Period: 1982Q1 - 1987Q4						
Worst Fund	-2.60	-0.17**	-5.83	-3.48*	-4.43	-3.16*
Best Fund	3.18	0.33	1.16	1.05	2.09	1.96
Best - Worst	5.78	0.53	6.99	3.42**	6.51	3.37**

* p-value is below 5%; ** p-value is below 1%.

[a]EWMF is the equally-weighted portfolio of all mutual funds in our sample.

[b]The notation mE nH indicates an m-quarter evaluation period and an n-quarter holding period. For example, 4E 2H indicates a portfolio for which the fund selections are based on an evaluation interval of the most recent 4 quarters and the selections are updated every 2 quarters.

[c]The mean returns over the whole sample for the S&P 500 and the EWMF were 1.58 and 1.88, respectively. For the first half, these mean returns were 0.59 and 2.30, respectively; and for the second half, 2.56 and 1.46.

[d]Sharpe's measure is the mean of the quarterly returns divided by the standard deviation of the quarterly returns. The Sharpe's measures over the whole sample period for the S&P 500 and the EWMF were 0.19 and 0.20, respectively (First half: 0.09 and 0.27, Second half: 0.26 and 0.14). For each of the Sharpe's measures, a z-statistic was calculated to test the significance of the difference in Sharpe's measure between the strategy and the EWMF. Significance of the z-statistic is indicated by asterisks.

[e]The corrected estimates take into account fund-specific heteroscedasticity. Since the best and worst fund portfolios are constructed from one fund each year, the residual variance of the portfolios for that year will be proportional to the residual variances of the funds which make up the portfolios in that year. The residual variances for the individual funds were calculated from the regressions of the fund returns on the benchmark series over the entire series. The corrected estimates are then calculated via weighted least squares where the weights for each year are proportional to the residual variances of the funds included in the portfolios for that year.

Table 6

Timing Versus Selectivity of Hot Hands Strategy

Quarterly Returns: 1976Q1 - 1987Q4

	4E4H result	Selected Percentiles from 5000 Draws[a] 5%	10%	50%	90%	95%
Mean excess return (%)	4.33	1.60	1.82	2.67	3.54	3.77
Sharpe's measure	0.35	0.14	0.17	0.24	0.32	0.35
Jensen's alpha (%)	2.45	-0.12	0.11	0.93	1.78	2.02

[a]Each draw generates a portfolio that includes, each year, two funds randomly chosen from among the 19 that appear in a 4E4H strategy that selected the best two funds based on Jensen's α. The probability of including a particular fund in any year is set equal to the fund's relative frequency observed in the 4E4H strategy.

Table 7

Performance of 4E4H Strategy Relative to Benchmark of P8 Portfolios

Quarterly Returns: 1976Q1 - 1984Q4

Reference Benchmark	Jensen's α	t-stat (Null: $\alpha = 0$)	$\bar{R}$-squared[a]
Best Fund Selected			
SP500[b]	3.37	2.35*	0.56
P8[c] (eight portfolios)	3.21	1.94**	0.77
Best Two Funds Selected			
SP500	2.38	2.12*	0.72
P8 (eight portfolios)	2.97	2.27*	0.85
SP500[d] (1975Q1 - 1984Q4)			
P8 (eight portfolios)	0.06	0.21	0.98

*p-value is below 5% **p-value is below 10%

[a]The $\bar{R}^2$ value is from the regression, which estimates Jensen's α, of the strategy's returns on the returns of the benchmark portfolio(s).

[b]The SP500 benchmark mimics the Standard & Poor's 500 index, with dividend reinvestment. The results with SP500 in this table differ from table 5 because the selection criterion is Jensen's α rather than net return. The conclusions are the same with either selection criterion.

[c]The P8 benchmark comprises eight portfolios chosen to account for anomalies related to firm size, dividend yield, and past returns — see Grinblatt and Titman (1987).

[d]The results of SP500 versus P8 are only included to show that the SP500 benchmark is not mispriced relative to P8.

Table A1

Basic Statistics

Quarterly Returns from 1974:4 - 1988:1

Panel A: Summary Statistics on Benchmark Portfolios

Benchmark	Mean (%)	Std. Dev. (%)
3-month Treasury Bills	1.99	0.70
Standard & Poor's 500 (SP500)		
total returns	4.18	8.67
excess return	2.19	8.80
Portfolio of all mutual funds (EWMF)		
total returns	4.45	9.60
excess returns	2.46	9.74

Panel B: Market Model Estimates

Regression: $R_{it} - R_{ft} = \alpha_i + \beta_i(R_{mt} - R_{ft}) + \epsilon_{it}$; benchmark portfolio, m, is SP500.

Fund	α	t-stat (Null: $\alpha = 0$)	β	t-stat (Null: $\beta = 1$)	Q-stat[a]	(p-value)
EWMF (Portfolio of all mutual funds)	0.12	0.34	1.07	1.74	42.60	(0.00)
ACORN FUND	1.48	2.16	1.08	1.07	18.76	(0.09)
AFUTURE FUND	-0.42	-0.55	1.03	0.39	61.94	(0.00)
AMERICAN INVESTORS FUND	-1.56	-1.40	1.34	2.73	22.10	(0.04)
BABSON GROWTH FUND	-0.73	-2.48	1.03	0.79	9.97	(0.62)
BEACON HILL MUTUAL FUND	-0.75	-1.81	0.86	-3.09	35.84	(0.00)
BOSTON CO. CAP. APPREC.	-0.17	-0.40	0.95	-1.13	20.01	(0.07)
BULL & BEAR CAP GROWTH	-0.05	-0.08	1.24	3.16	22.19	(0.04)
BULL & BEAR EQUITY INC.	0.07	0.18	0.74	-5.61	7.09	(0.85)
CENTURY SHARES TRUST	0.47	0.51	0.90	-0.98	13.64	(0.32)
CHARTER FUND	1.02	1.59	0.92	-1.11	32.36	(0.00)
COLUMBIA GROWTH FUND	0.69	1.29	1.14	2.27	13.23	(0.35)
COMPANION FUND	-0.17	-0.63	1.06	1.81	14.16	(0.29)
COMPOSITE FUND	0.08	0.14	0.82	-2.80	14.97	(0.24)
CONCORD FUND	0.08	0.11	0.86	-1.65	5.60	(0.94)
CONSTELLATION GROWTH	0.44	0.44	1.57	5.15	13.64	(0.32)
DODGE & COX STOCK	0.48	2.43	0.96	-1.82	9.44	(0.67)
DREYFUS THIRD CENTURY	0.79	1.03	0.97	-0.34	16.06	(0.19)
ELFUN TRUSTS	0.46	1.65	0.99	-0.19	15.63	(0.21)
ENERGY FUND	0.41	0.70	0.83	-2.58	16.74	(0.16)
EVERGREEN FUND	2.30	2.71	1.24	2.55	26.63	(0.01)

TABLE A1 ... continued

Panel B (continued)

Fund	α	t-stat (Null: $\alpha = 0$)	β	t-stat (Null: $\beta = 1$)	Q-stat[a]	(p-value)
EXPLORER FUND	-1.14	-0.80	1.26	1.62	30.03	(0.00)
FIDELITY CONTRAFUND	-0.07	-0.14	1.12	1.98	8.42	(0.75)
FIDELITY DESTINY	1.74	2.77	1.19	2.70	9.09	(0.69)
FIDELITY FUND	0.20	0.75	0.98	-0.79	8.34	(0.76)
FIDELITY TREND	-0.42	-0.97	1.14	2.83	11.02	(0.53)
FINANCIAL DYNAMICS FUND	0.05	0.07	1.15	1.85	16.17	(0.18)
FINANCIAL INDUSTRY FUND	0.18	0.40	0.96	-0.69	38.05	(0.00)
FORTY-FOUR WALL ST FUND	-2.62	-1.20	1.97	4.00	53.54	(0.00)
FOUNDERS GROWTH FUND	-0.06	-0.12	1.03	0.46	22.41	(0.03)
FOUNDERS MUTUAL FUND	-0.86	-2.58	1.04	1.00	11.50	(0.49)
FOUNDERS SPECIAL	-0.20	-0.22	1.03	0.25	12.26	(0.42)
GENERAL SECURITIES	0.46	0.65	0.92	-1.04	3.73	(0.99)
GROWTH INDUSTRY SHARES	0.11	0.20	1.08	1.23	13.23	(0.35)
GUARDIAN MUTUAL FUND	0.81	2.14	0.94	-1.43	10.44	(0.58)
HARTWELL GROWTH FUND	0.89	0.96	1.29	2.83	12.69	(0.39)
HARTWELL LEVERAGE FUND	0.48	0.38	1.65	4.61	11.81	(0.46)
HORACE MANN GROWTH FUND	-0.55	-1.37	1.04	0.98	11.76	(0.47)
INDUSTRY FUND OF AMERICA	-1.13	-0.93	1.19	1.39	29.72	(0.00)
IVY GROWTH FUND	0.45	1.10	0.86	-3.04	22.01	(0.04)
JANUS FUND	0.68	0.86	0.90	-1.18	14.30	(0.28)
KEYSTONE INTL FUND	0.34	0.51	0.90	-1.40	6.71	(0.88)
KEYSTONE K-2	-0.56	-1.43	1.03	0.73	31.38	(0.00)
KEYSTONE S-1	-1.03	-2.95	1.03	0.67	13.14	(0.36)
KEYSTONE S-3	-0.17	-0.29	1.23	3.64	31.21	(0.00)
KEYSTONE S-4	-0.45	-0.53	1.50	5.36	25.23	(0.01)
LEHMAN INVESTORS FD, INC	0.02	0.07	0.96	-1.15	14.26	(0.28)
LEXINGTON GROWTH FUND	0.07	0.09	1.25	3.00	27.09	(0.01)
LEXINGTON RESEARCH FUND	0.00	0.01	0.95	-1.02	19.57	(0.08)
LOOMIS-SAYLES CAP. DEV.	1.15	1.41	1.21	2.35	13.66	(0.32)
MANHATTAN FUND	-0.07	-0.18	1.10	2.33	17.90	(0.12)
MATHERS FUND	1.35	1.73	1.00	-0.03	11.71	(0.47)
MEESCHAERT CAP. ACCUM.	-0.52	-1.12	0.69	-6.06	4.55	(0.97)
MORGAN (W.L.) GROWTH	0.26	0.55	1.13	2.48	14.14	(0.29)
MUTUAL SHARES CORP.	2.12	3.47	0.73	-3.90	9.93	(0.62)
NATIONAL INDUSTRIES FUND	-0.67	-1.41	0.93	-1.32	49.74	(0.00)
NEUWIRTH FUND	-0.63	-0.84	1.24	2.88	10.12	(0.61)
NEWTON GROWTH FUND	-0.41	-0.59	1.12	1.55	11.63	(0.47)
NICHOLAS FUND	1.67	2.51	0.99	-0.15	14.33	(0.28)
OMEGA FUND	-0.17	-0.18	1.08	0.82	13.82	(0.31)
ONE HUNDRED FUND	-0.52	-0.54	0.95	-0.46	9.86	(0.63)

TABLE A1 ... continued

Panel B (continued)

Fund	α	t-stat (Null: $\alpha = 0$)	β	t-stat (Null: $\beta = 1$)	Q-stat[a]	(p-value)
ONE HUNDRED ONE FUND	-0.29	-0.48	0.86	-2.07	14.39	(0.28)
PARTNERS FUND	1.27	2.98	0.71	-6.21	9.40	(0.67)
PENN SQUARE MUTUAL	0.16	0.45	0.98	-0.54	17.23	(0.14)
PENNSYLVANIA MUTUAL FUND	2.11	1.67	1.33	2.33	6.17	(0.91)
PINE STREET FUND	-0.21	-0.77	0.92	-2.52	8.08	(0.78)
PRICE (ROWE) GROWTH STK.	-0.87	-2.39	1.05	1.22	13.44	(0.34)
PRICE (ROWE) NEW ERA	0.16	0.27	1.05	0.73	36.51	(0.00)
PRICE (ROWE) NEW HORIZ.	-0.25	-0.34	1.29	3.53	28.61	(0.01)
RAINBOW FUND	-0.54	-0.62	0.89	-1.15	7.52	(0.82)
SAFECO EQUITY FUND	0.27	0.58	1.07	1.41	18.74	(0.10)
SAFECO GROWTH FUND	0.83	1.17	1.13	1.62	24.07	(0.02)
SCUDDER COMMON STOCK	0.02	0.05	0.93	-1.84	16.13	(0.19)
SCUDDER DEVELOPMENT FUND	0.70	0.66	1.30	2.56	14.13	(0.29)
SCUDDER INTERNATIONAL FD	0.63	0.79	0.82	-2.01	11.02	(0.53)
SELECTED SPECIAL SHARES	-0.72	-1.27	1.03	0.46	7.72	(0.81)
SEQUOIA FUND	2.40	2.66	0.77	-2.27	11.87	(0.46)
SHERMAN, DEAN FUND	-0.70	-0.33	1.09	0.41	16.67	(0.16)
SMITH, BARNEY EQUITY	0.14	0.31	0.92	-1.59	10.36	(0.58)
STATE FARM GROWTH FUND	0.75	1.51	0.99	-0.22	60.12	(0.00)
STATE STREET INV. CORP.	-0.01	-0.01	1.02	0.39	21.96	(0.04)
STEADMAN AMERICAN INDUS.	-3.24	-4.21	0.98	-0.27	19.96	(0.07)
STEADMAN INVESTMENT	-2.08	-3.21	0.85	-2.06	5.39	(0.94)
STEADMAN OCEANOGRAPHIC	-3.24	-2.93	1.02	0.17	17.10	(0.15)
STEIN R&F CAPITAL OPPORTUNITY	0.21	0.24	1.32	3.20	25.03	(0.01)
STEIN R&F STOCK	-0.59	-1.14	1.20	3.38	14.22	(0.29)
STRATTON GROWTH FUND	-0.10	-0.16	1.10	1.44	6.10	(0.91)
TUDOR FUND	0.81	1.07	1.17	2.00	10.31	(0.59)
TWENTIETH CENTURY GROWTH	2.17	1.89	1.45	3.49	34.48	(0.00)
TWENTIETH CENTURY SELECT	2.21	2.68	1.21	2.29	24.21	(0.02)
UNIFIED MUTUAL SHARES	-0.18	-0.49	0.88	-2.95	10.96	(0.53)
USAA MUTUAL FD GROWTH	-1.01	-2.14	1.12	2.29	6.24	(0.90)
VALUE LINE FUND	0.55	0.63	1.13	1.37	16.78	(0.16)
VALUE LINE LEVER. GROWTH	1.68	1.56	1.19	1.58	14.05	(0.30)
VALUE LINE SPECIAL SITUATION	0.02	0.02	1.40	3.23	30.28	(0.00)
WEINGARTEN EQUITY FUND	1.54	1.95	1.28	3.24	20.40	(0.06)
WINDSOR FUND	1.56	2.96	0.93	-1.28	7.33	(0.83)

[a]The Q-statistic is based on the first twelve autocorrelations of the residuals from the market model regressio see equation (4). Under the null hypothesis that the autocorrelations are zero, the Q-statistic is distributed asympt cally χ^2_{12}. The p-value, i.e., the probability of the observed Q-statistic under the null hypothesis, is shown in parantheses.

Table A2

Joint Tests of Zero Jensen's Alpha

Quarterly Returns: 1974Q4 - 1988Q1

SUR Approach 2: Randomly select a subset of 40 funds from 96 available funds and estimate full cross-sectional residual covariance matrix. Repeat selection 100 times. Under the joint null hypothesis of zero Jensen's alpha for each included fund, the g-statistic has a 1% nominal significance level of 64.

	Percentiles from 100 Draws				
	min	5%	50%	95%	max
g-statistic	231	274	580	1255	1874

SUR Approach 3: Estimate a full covariance matrix for residuals of 96 funds. Under the joint null hypothesis of zero Jensen's alpha for each included fund, the g-statistic has a 1% nominal significance level of 130.

A. Assume an average cross-correlation. (The average cross-correlation for the sample is 0.23.)

	Assumed Cross-Correlation				
	+0.5	+0.23	0.00	-0.25	-0.5
g-statistic	432	281	219	173	144

B. Assume that the residual correlations are adequately represented by a five-factor model.

g-statistic = 496

Note: For SUR Approach 2, we also studied a modification to the g-statistic that makes it analogous to a F-statistic, which is used to test restrictions on an equation-by-equation basis. We adjusted the degrees of freedom to account for the estimation of the covariance matrix. The F-type statistic could have superior finite sample properties and a test based on it is more conservative than the one based on the g-statistic. Fortunately, in our case, the inferences stay unchanged: the values that we obtain for the F-type statistic — not reported — have p-values below 0.1%.

Figure 1
Mean Annual Returns Percentiles: Year t vs. Year t+1

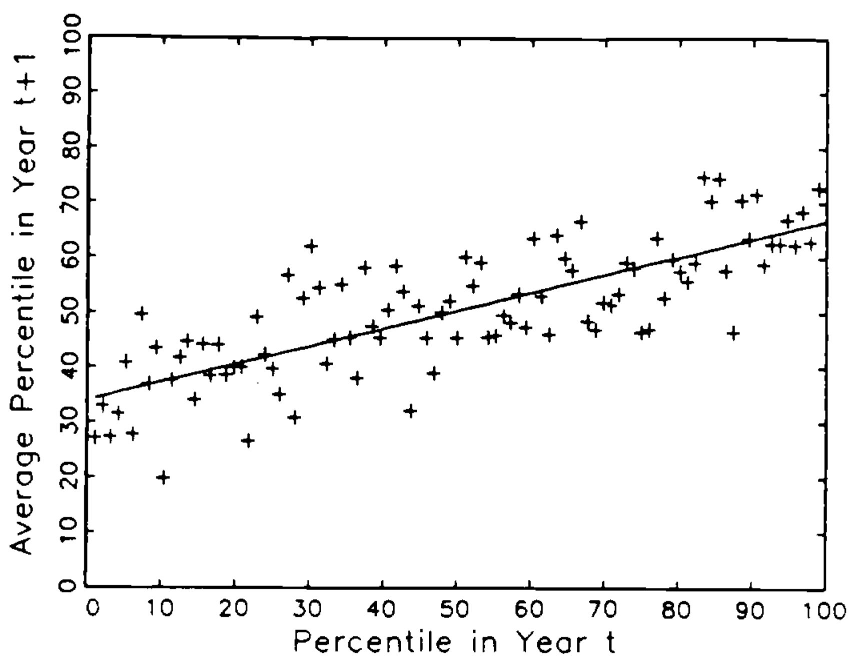

Figure 2
Decile Rank in Year t+1 for Best Fund in Year t