Hot Hands in Mutual Funds: Short-Run Persistence of Relative Performance, 1974-1988

Darryll Hendricks, Jayendu Patel, Richard Zeckhauser

Journal of Finance, Volume 48, Issue 1 (Mar., 1993), 93-130.

Your use of the JSTOR database indicates your acceptance of JSTOR's Terms and Conditions of Use. A copy of JSTOR's Terms and Conditions of Use is available at http://www.jstor.org/about/terms.html, by contacting JSTOR at jstor-info@umich.edu, or by calling JSTOR at (888)388-3574, (734)998-9101 or (FAX) (734)998-9113. No part of a JSTOR transmission may be copied, downloaded, stored, further transmitted, transferred, distributed, altered, or otherwise used, in any form or by any means, except: (1) one stored electronic and one paper copy of any article solely for your personal, non-commercial use, or (2) with prior written permission of JSTOR and the publisher of the article or other text.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Journal of Finance is published by American Finance Association. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/afina.html.

Journal of Finance
©1993 American Finance Association

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2001 JSTOR

http://www.jstor.org/
Mon Aug 13 10:40:35 2001

THE JOURNAL OF FINANCE • VOL. XLVIII, NO. 1 • MARCH 1993

Hot Hands in Mutual Funds: Short-Run Persistence of Relative Performance, 1974–1988

DARRYLL HENDRICKS, JAYENDU PATEL,
and RICHARD ZECKHAUSER*

ABSTRACT

The relative performance of no-load, growth-oriented mutual funds persists in the near term, with the strongest evidence for a one-year evaluation horizon. Portfolios of recent poor performers do significantly worse than standard benchmarks; those of recent top performers do better, though not significantly so. The difference in risk-adjusted performance between the top and bottom octile portfolios is six to eight percent per year. These results are not attributable to known anomalies or survivorship bias. Investigations with a different (previously used) data set and with some post-1988 data confirm the finding of persistence.

ACADEMIC STUDIES SINCE THE 1960s find that mutual funds do not systematically outperform benchmark portfolios (such as the "market" indices)—see the classic papers by Treynor (1965), Sharpe (1966), and Jensen (1968), and recent updates with refinements by Grinblatt and Titman (1989b), Connor and Korajczyk (1991), and references therein. The practitioner literature sees matters differently, expressing a consistent belief that active selection among actively managed funds can be profitable. 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, p. 14) 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."

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 do not, search for

*John F. Kennedy School of Government, Harvard University. 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 interest in the hot hands phenomenon stimulated this investigation, Sheridan Titman for providing data on the P8 indices, and David Modest for allowing us to use the Henriksson-Lehmann-Modest sample of mutual fund returns. Helpful comments from Richard Ruback, Lawrence Summers, G. William Schwert, two referees, the editor René Stulz, and seminar participants at Cornell University, the conference on "Economics of Financial Markets" at Johns Hopkins University, the Harvard Business School, the Kennedy School of Government, and the National Bureau of Economic Research are gratefully acknowledged.

the measure that puts them in the best possible light. Directly or indirectly, investors are willing to act on such information of relative performance: Patel, Zeckhauser, and Hendricks (1992) document that investors steer their money to funds that have performed well recently. Are such investors mistaken?

Grinblatt and Titman (1989b) report some statistical evidence of persistence in mutual fund returns over five-year periods, but are unable to realize economically significant strategies based on this pattern. Jensen (1968) had concluded similarly in his earlier classic study.[1] We reassess the extent to which the relative performance of mutual funds can be reliably predicted by studying shorter evaluation horizons. We examine quarterly returns data over 1974-88 on a sample of open-end, no-load, growth-oriented, equity funds constructed to mitigate survivorship bias. Statistical evidence on performance is computed in relation to fellow funds as well as to a variety of market indices. The persistence of relatively superior fund performance proves to be significant, although it is predominantly a short-run phenomenon, peaking at roughly four quarters. Adopting the argot of the sports world, we say that funds delivering sustained short-run superior performance have "hot hands."[2] We can also identify ex ante underperformers with substantial negative excess returns. The ability to predict reliably the octile rank of the funds in our sample is robust across all the short-run evaluation periods we consider (from one to eight quarters). The strongest results appear when the evaluation period is one year, which is the lag-length beyond which the partial autocorrelations in excess returns become insignificantly different from zero. We establish that ex ante investment strategies based on identifying funds with hot (and icy) hands successfully sort among funds, and can improve risk-adjusted returns by 6% per year. Superior performance vis-à-vis traditional benchmarks offers an excess return of 3 to 4% excess per year.

The results are robust to several explorations that we pursue: (1) assessment with a variety of benchmark portfolios, including a multi-portfolio benchmark that accommodates known anomalies such as firm size effects and returns reversions; (2) simulations that evaluate whether the hot hands

[1]After we completed two revisions of our research, independent and parallel research by Goetzmann and Ibbotson (1990, Working paper, School of Management, Yale University) was brought to our attention. Similar to this paper, Goetzmann and Ibbotson conclude in favor of short-term persistence in returns of mutual funds. However our study differs substantially from Goetzmann and Ibbotson in analytic approaches, checks for robustness, and construction of mutual fund samples.

In other research, Ippolito (1989) reports that fund managers do sufficiently better than passive strategies to cover their fees and loads, a finding that suggests that fund managers do uncover mispriced assets even though no significant deviations from efficiency remain for uninformed investors. However, Elton, Gruber, Das, and Hlavka (1991) suggest that Ippolito's finding is not robust.

[2]Camerer (1989) examines point spreads for betting on professional basketball games. He finds that bettors respond too strongly to winning and losing streaks, a finding consistent with having them believe in a "hot hand" phenomenon for teams. Unfortunately, bookmakers' commissions preclude profiting from Camerer's findings.

strategy merely picks the good performers in the sample or rather picks the right fund at the right time; (3) adjustments for nonlinearities/option-like characteristics; (4) allowance for time variation in betas due to the changing composition of included funds in the hot hands portfolios; and (5) confirmation with a different data set of mutual funds, which has been previously used in the literature, as well as with out-of-sample evidence from 1989–90. Recently, Brown *et al.* (1992) argue that studies of performance persistence among mutual funds are contaminated by survivorship bias. Our investigation reveals that such bias is not a material problem for our study.

The remainder of the paper is organized as follows. Section I presents the statistical results that identify short-run persistence of performance. Section II demonstrates that short-run performance persistence in mutual funds can be exploited to achieve economically valuable investment strategies, and that these claims are robust. Section III provides out-of-sample evidence, and addresses concerns about selection bias, particularly survivorship bias. Section IV concludes, noting that substantial gains are available from investing in recent mutual fund winners. Neither traditional anomalies nor survivorship bias explains these gains.

I. Statistical Autocorrelation in Mutual Fund Performance

A null hypothesis of an efficient market for mutual funds and martingale equity prices[3] implies that historical performance cannot be used to identify mutual funds that will be superior performers in the future. We first establish that relative mutual fund excess returns exhibit serial correlation. In later sections, we show that such statistical autocorrelation lends itself to economically worthwhile fund selection strategies.

A. Hypotheses

Let r_{it} be the net return in excess of the risk-free rate (proxied by Treasury bill yields in our empirical work) of mutual fund i in quarter t. We can decompose the mutual fund excess return as follows:

$$r_{it} = M_{t-1}(r_{it}) + \varepsilon_{it}. \tag{1}$$

Here $M_{t-1}(\cdot)$ denotes the market's expected value that is conditioned on the information available to the market at $t-1$, say I_{t-1}; ε_{it} is the residual return realized in quarter t.

The usual null hypothesis (of market efficiency/rational expectations) is that ε is unpredictable:

HYPOTHESIS 1: $E_{t-1}(\varepsilon_{it}) = 0$, *for all i and t.*

[3]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 and Stock (1989) suggest that the evidence for mean reversion may be overstated.

Here $E(\cdot)$ denotes the mathematical expectation; the $t-1$ subscript indicates that the expectation is conditioned on the information I_{t-1}. Since I_{t-1} includes at least the past history of fund returns, Hypothesis 1 implies, for instance, that $E(\varepsilon_{it}\varepsilon_{it-k}) = 0$ for $k > 0$. The typical alternate hypothesis is that some funds have a constant nonzero ex ante excess performance, perhaps due to the ability of the fund managers to identify underpriced equities or to time market movements:

HYPOTHESIS 2: $E(\varepsilon_{it}) > 0$ *for some* i.

An alternate hypothesis, which allows for the unconditional mean of ε to be zero, admits short-run predictability of residual returns from past history (violating a weak-form market efficiency):

HYPOTHESIS 3: $E(\varepsilon_{it}|\varepsilon_{it-1}, \varepsilon_{it-2}, \varepsilon_{it-3,\ldots}) \neq 0$ *for some* i *and some* t.

For discussion of our cross-sectional time-series tests, we specialize Hypothesis 3 to a linearized projection form:

HYPOTHESIS 3′: $E(\varepsilon_{it}|\varepsilon_{it-1}, \varepsilon_{it-2}, \varepsilon_{it-3,\ldots}) = \sum_{j=1}^{J} \rho_{ij}\varepsilon_{it-j}$; $\rho_{ij} \neq 0$ *for some* i *and* j.

Although most studies examine for nonzero autocorrelation in residual returns (Hypothesis 3′) merely as a routine specification check, we explore it in depth. If the ρ's are positive, then Hypothesis 3′ implies that funds have hot (and icy) hands—that is, funds' recent relative performance will persist, at least in the near future. It is important to note that the Hypothesis 3′ can be studied independently of Hypothesis 2, although many tests might be unable to distinguish between them. For example, tests that regress estimated performance measures in one period on estimated performance measures in past periods could conceivably be taken as evidence on either or both hypotheses, depending on the performance measure used.

B. Sample Construction: Criteria, Characteristics, and Sources

We study mutual fund performance, based on quarterly returns that are net of management fees, over the period 1974 to 1988. All dividends are assumed to be reinvested by purchase of shares in the mutual fund at the end of the quarter in which the dividends are distributed. We transform all returns to excess returns by subtracting the returns on one-month U.S. Treasury bills (whose source is Ibbotson Associates (1991)) over the quarter.

Prior to 1982, we include mutual funds that meet the following two criteria. First the fund had at least a four-quarter price and dividend history available at the end-of-previous-year issue of *Wiesenberger Mutual Funds Current Dividend and Price Record*. Second, the fund is listed in the most recent *Wiesenberger Mutual Funds Panorama* as a no-load fund, open to all investors, and as pursuing a growth, aggressive growth, or growth and income objective. We start selecting funds from the first quarter of 1975 (whose data

is collected from the first quarter of 1974), and stop adding new funds after the fourth quarter of 1982. Once admitted into the sample, the four-quarter return history (or a longer history if possible) is combined with all valid data through 1988.

The returns on the eligible funds in each quarter are recorded from four sources:

1. *Wiesenberger Mutual Funds Current Dividend and Price Record* (1974–88);
2. CDA Investment Technologies, Silver Spring, Md. (1975–84);
3. Barron's publication of data collected by Lipper Analytical Services (1982–88); and
4. Higgins Associates, Cambridge, Mass. (1989–90, for some out-of-sample performance).

The last three sources provide calculated returns, while the first provides only price and distribution data. The source chosen for each time period depends on availability, and its ease of producing machine-readable data. Since our criterion of sample inclusion relies on the first source, we are assured of obtaining data on all funds in the sample.[4]

The total number of funds in our sample is 165, though the number of funds in any quarter varies. In 1982, the date after which we stopped adding new funds, our sample includes 121 funds. By the end of 1988, our sample shrinks to 94 funds; these are the funds we track through the out-of-sample 1989–90 period.

We restrict our sample to no-load funds because the transactions costs associated with investing in (and switching between) such funds are close to zero, which is convenient for evaluating the switching strategies we consider. (We ignore tax consequences.) Further, we restrict our sample to growth funds to secure a homogeneity in objectives and institutional characteristics that makes it reasonable to presume similar equilibrium expected returns across the funds. Predictable differences in fund performance across our sample are thus less likely to be attributable to uncontrolled variations in equilibrium expected returns.

We attempt to minimize the impact of survivorship bias in our sample. The data record on a fund in our sample ends only if it ceases to exist, changes to a non-growth objective, merges into another fund, closes to new investors, adopts a load fee, or becomes a closed-end fund. In all such cases, returns data are recorded through the quarter immediately preceding the data-terminating event.[5] If there are suprise terminations within a quarter, this would introduce some survivorship bias in our sample. Fortunately, because

[4]Providers of mutual fund returns data, including the last three of our sources, calculate returns using the same price and distribution data that are published in the *Wiesenberger Mutual Funds Current Dividend and Price Record*. The discrepancies between data sources are virtually nil, a point we confirmed by examining fund returns in periods where sources overlap.

[5]Since information on funds that became restricted or adopted a load fee are only available annually, we retained such funds in the sample through end-of-year of the event.

of the need to obtain the approval of fund shareholders via a vote, most of these events must be publicly disclosed more than a quarter in advance. This implies that every investor has sufficient notice to close out his position in the fund at the end-of-quarter preceding the quarter in which the fund entity no longer meets our criteria. Such advance knowledge of fund termination permits us to examine strategies based on full quarter returns without risk of inappropriately relying on information that an investor could not have had. In Section III below, we directly address concerns of survivorship bias, emphasized recently by Brown *et al.* (1992), and conclude that inferences about persistence performance with our sample are robust to this problem.

Our exclusive focus on equity funds makes reasonable our reliance on equity portfolio benchmarks, with dividend reinvestment.[6] However, simple indices used as equity benchmarks during the 1970s are now well-known to be mean-variance inefficient with respect to passive strategies such as those based on firm size and dividend yields—see De Bondt and Thaler (1989), Fama (1991), and references therein. We provide an agnostic evaluation using the following three benchmark sets:

1. Single portfolio benchmarks that have been commonly used in performance and equity pricing studies, i.e., an equally weighted index of equities on the NYSE (denoted EWNYSE), an equally weighted index of equities trading on the NYSE and the AMEX (denoted EWCRSP, since the index is taken from the CRSP tapes), and a value-weighted index of the equities trading on the NYSE and the AMEX (denoted VWCRSP). Our results with VWCRSP are always very similar to those with the well-known Standard & Poor's 500 index, and therefore results with the latter are not reported in the interest of parsimony.
2. An eight-portfolio benchmark, P8. P8 is constructed by Grinblatt and Titman (1989b) to account for anomalies related to firm size, dividend yield, and mean reversion in equity returns. The P8 benchmark is available for the 1975–84 subperiod.
3. The equally weighted index of mutual funds in our sample, EWMF. If the characteristics of funds included in our sample are sufficiently similar (especially in terms of their loading on market-relevant priced factors), as seems plausible given our stringent fund inclusion criteria, EWMF will serve as a good composite benchmark. Since we use a linear regression of the returns on individual funds on EWMF to adjust for risk, we will need to assume that the relation be linear in the true underlying factors (such as in the Arbitrage Pricing Theory). To the extent that some priced factors are unknown to the econometrician or not agreed upon by the finance profession, the EWMF may provide a better benchmark than the others discussed above.

[6]Generally, investors can reap substantial diversification gains by including other assets besides equities in their portfolios, such as real estate (Firstenberg, Ross, and Zisler (1988)) or foreign securities. Thus, even the multiple portfolio benchmarks we consider, which remain based on U.S. equities, are unlikely to be globally efficient.

In the Appendix, we provide summary statistics on the benchmarks and on the individual funds in our sample. Briefly, the mutual fund betas (versus the VWCRSP) are distributed around unity (median of 0.99, interquartile range of 0.25), and most of the estimated individual α's (excess returns) are not significantly different from zero. Implementable investment strategies—not reported—that implicitly rely on some funds having a constant positive Jensen's α fail to generate significant ex ante excess returns (either statistical or economic); this finding is similar to that of Grinblatt and Titman (1989b, Table 9).

C. Short-Run Performance Persistence

We investigate for predictable residual returns, even though they may be unconditionally zero, in our sample. (The existence of funds with such patterns, of course, need not imply that economically worthwhile investment strategies are available. We defer to Section II the assessment of economic significance.) The unbalanced panel of funds across quarters and the likely nonzero cross-correlations among residual returns in any quarter preclude a direct test for zero serial correlation.

We rely on an approach based on Fama and Macbeth (1973), which is implemented, most directly for our purposes, in Jegadeesh (1990). Consider a cross-sectional regression for quarter t with the following structure:

$$r_{it} - M_{t-1}(r_{it}) = k_t + \sum_{j=1}^{J} a_{jt} r_{it-j} + u_{it}; \; i = 1, \ldots, N_t. \tag{2}$$

Here we have complete observations on N_t funds in periods $t - J$ through t. Under the null hypothesis (Hypothesis 1 or Hypothesis 2), the a's should be zero. Under the alternative of a nonzero serial correlation in the individual funds (nonzero ρ's in Hypothesis 3), the a's will be nonzero (unless there are unlikely exact cancellations). Since, under the null hypothesis, the a-estimates from different quarters are independent, a t-statistic can be computed for each of the J sets of $\{\hat{a}_{j1}, \hat{a}_{j2}, \ldots, \hat{a}_{jT}\}$.

Note that the test of persistence using stacks of cross-sectional regressions will not be subject to the survivorship bias problems discussed in Brown *et al.* (1992). Brown *et al.* point out difficulties in testing for relative performance persistence using methods that are adversely affected by truncation-induced differences in conditional fund performance means. The tests underlying Table I are free from such problems, however, since the time-series mean of each fund's performance measure is zero by construction.

To operationalize such t-tests based on estimates of (2), we need to specify $M_{t-1}(r_{it})$, the market equilibrium return. One simple approach assumes that $M_{t-1}(r_{it})$ is constant over the sample period. The results with demeaned r_{it} are shown in the first row of Table I. The t-statistics (greater than 2.5 at lags 2 and 4) indicate that the null hypothesis of non-autocorrelated residual returns can be rejected. The a-coefficients for the first four lags are all

Table I

Persistence Estimates from Stacked Cross-Sectional Regressions

For each quarter in 1976 through 1988, cross-sectional regressions are estimated by OLS where the dependent variable is a measure of the "residual" performance of fund i in the quarter. The independent variables consist of lags 1 to 8 of the dependent variable. The OLS-reported t-statistics (not shown) from the individual cross-sectional regressions are biased because of the correlation across fund returns; however, the estimated slope coefficients are unbiased. The set of slope coefficients from the different quarterly regressions is treated as a random sample from a normal population, and tests are performed.

Dependent Variable[a]	Time-Average of Cross-Sectional Coefficients (t-Statistics in Parentheses)									F-Test of Zero Coefficients[b]	
	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{a}_4$	$\hat{a}_5$	$\hat{a}_6$	$\hat{a}_7$	$\hat{a}_8$	$\Sigma\hat{a}_i$	Lags 1 to 4	Lags 5 to 8
$r_{it} - \bar{r}_t$	0.05	0.09	0.06	0.11	−0.05	−0.06	−0.06	0.06	0.20	5.02	2.24
	(1.16)	(2.55)	(1.52)	(2.89)	(−1.42)	(−1.99)	(−1.85)	(2.25)		[0.00]	[0.08]
Market model residual	0.05	0.10	0.06	0.11	−0.05	−0.05	−0.06	0.04	0.18	4.52	1.95
using EWNYSE	(1.03)	(2.56)	(1.62)	(2.88)	(−1.69)	(−1.84)	(−1.79)	(1.61)		[0.00]	[0.12]
Market model residual	0.04	0.10	0.07	0.10	−0.04	−0.05	−0.05	0.03	0.20	4.79	1.81
using VWCRSP	(0.94)	(2.63)	(2.01)	(2.90)	(−1.51)	(−1.92)	(−1.57)	(1.46)		[0.00]	[0.14]
Market model residual	0.03	0.09	0.06	0.10	−0.04	−0.05	−0.05	0.03	0.17	4.35	2.00
using EWMF	(0.74)	(2.33)	(1.85)	(2.93)	(−1.48)	(−2.16)	(−1.72)	(1.08)		[0.00]	[0.11]

[a]For the first row of results in the table, the measure of "residual" return for fund i is the excess return of fund i in quarter t less its mean over the entire sample period. For the remaining rows, the measure of "residual" return is the residual from a preliminary regression of the excess returns of fund i on the excess returns of the benchmark portfolio indicated in parenthesis (EWNYSE = Equally weighted NYSE index, VWCRSP = Value-weighted CRSP index containing both NYSE and AMEX shares, EWMF = Equally weighted portfolio of all mutual funds in the sample). In findings not reported, similar results are obtained with the residuals from multiportfolio benchmarks.

[b]The F-statistic tests the joint hypothesis that the slope coefficients for lags 1 to 4 (or lags 5 to 8) are zero. The F-statistic is computed as $\left(\frac{N(N-4)}{4(N-1)}\right)\hat{a}'\hat{\Sigma}^{-1}\hat{a}$, N (= 52) is the number of quarterly cross-sections, $\hat{a}$ is the 4×1 mean vector of persistence coefficients, and $\hat{\Sigma}$ is the sample covariance matrix of the coefficients. Under the null hypothesis, the statistic is distributed $F_{4,48}$ for a large number of funds per cross-section. The p-values based on the $F_{4,48}$ distribution are shown in brackets; the 5% (1%) critical value is 2.56 (3.74).

positive and jointly significant (the F-statistic that jointly tests if the t-statistics associated with lags 1 to 4 are zero has a p-value[7] below 1%); the a-coefficients from lags 5 to 8 are less clearly significant (the corresponding F-statistic has a p-value of 8%).[8]

We verify this pattern with alternative specifications of $M_{t-1}(r_{it})$ below. An alternative construction of the residual returns, the left-hand side of equation (2), is to posit a linear K-factor model for r_{it} (which allows possible time variation in $M_{t-1}(r_{it})$):

$$r_{it} = b_{0i} + \sum_{k=1}^{K} b_{ki} f_{kt} + \varepsilon_{it}. \tag{3}$$

Here the b's are the factor loadings on the K-factors whose realizations are the f's. In our implementation, we specify the factors to be benchmarks (which, for the purposes in this section, we assume to be correlated sufficiently with the latent factors to deliver $E_{t-1}(\varepsilon_{it}|r_{it-1}, r_{it-2}, \ldots, r_{it-L}) = 0$ under the null hypothesis). Equation (3) is, of course, nothing more than a multivariate "market" model. Rows two, three, and four of Table I report t-statistics where the proxy for the residual returns (left-hand side of equation (2)) is the residual from market model regressions (equation (3)) with different benchmark (factor) choices.

The results reject the hypothesis of no predictability in residual returns. The F-statistics, which test if the first four a-coefficients are zero, are extremely significant (p-values below 1%); as before, the F-statistics that test whether the partial autocorrelations at lags 5 through 8 are zero do not clearly reject the null hypothesis (p-values hover around 10%). Note that when the usual t-tests and F-tests are examining means, they are asymptotically consistent in the presence of limited heteroscedasticity satisfying White's (1980) conditions. Results from bootstrap simulations confirm that our conclusions are robust despite the small samples—see discussion and results in Table VII, Panel B.

[7]The p-value is the probability, under the null hypothesis, of obtaining a value larger than that actually computed.

[8]Examination of such coefficients is central to our analysis. It appears desirable to explore the robustness of the significance levels of associated tests with bootstrap simulations, especially since the sample sizes are quite limited. Unfortunately, bootstrap simulations cannot be conducted that preserve both the cross-sectional correlation structure and the relative frequency of a fund's availability for the hot hands strategy. That is because our sample includes funds that have data for only a portion of the time period under study. Thus, if we resample entire cross-sections randomly for our bootstrap in order to preserve cross-sectional correlations in performance, we will typically not have nine consecutive quarters of performance for many of the funds in the bootstrapped sample. We, therefore, limited the bootstrap investigation to a subsample of 51 funds that survived over the entire period. The results of this approach are discussed in detail in Section III surrounding Table VII. Briefly, the results suggest that the asymptotic significance levels reported in Table I are somewhat too conservative.

The pattern of the coefficients is such that there appears to be a positive performance persistence for four quarters and a reversal thereafter.[9] In the wake of a 1% superior performance, the cumulative residual gain is about 30 basis points over the next four quarters, but declines to about 20 basis points by eight quarters.

Of course, the persistence in relative performance seen in Table I could be due to a failure to correctly specify the structure of equilibrium returns. Other plausible conjectures for short-run persistence without long-run superior performers include:

1. superior analysts get bid away once they build a track record,
2. new funds flow excessively to successful performers, which then leads to a bloated organization and fewer good investment ideas per managed dollar,
3. urgency and drive are diminished once reputation is established,
4. market feel of managers is limited to evanescent market conditions, and
5. salaries and fees rise to capitalize on demands arising from recent successes.

In Table II, we assess the robustness of the statistical persistence documented in Table I. We only report results with the EWNYSE since results with other choices prove similar. In Panel A of Table II, we estimate the partial autocorrelations for the half samples. The coefficients are very similar between the half samples. A heteroscedastic-consistent χ^2-test, which is computed following Duncan's (1983) straightforward multivariate extension of White (1980),[10] does not reject the null hypothesis that the partial autocorrelations at each lag are equal between the half samples (p-value = 0.75).

In Panel B of Table II, we explore for seasonality by calendar quarter, an investigation motivated by the well-known January effect in equity returns. The short-run performance in persistence displays calendar-quarter seasonality: we can easily reject the null hypothesis that the coefficients at the different lags are equal across quarter subsamples—the p-value for an overall χ^2-test is zero. The seasonality that we observe is not a January (first

[9] An interesting feature of the pattern in the coefficient estimates is that the size of the coefficient for the first lag is smaller than that relative to lags 2 through 4. Based on bootstrap simulations, the significance levels for the a-estimates at the first lag were much lower. We have no simple explanation for the difference in the size of the estimates relative to lags 2 through 4. One possibility is the confounding of the persistence effect (found clearly in lags 2 through 4) by the general one-month returns reversal in equities (Jegadeesh (1990)). Alternatively, if the true persistence coefficients, though positive and declining with lag for each fund, vary across funds, then the a-coefficients for the *cross-sectional* equation (2) need not display the same declining relation with lag length.

[10] We thank a referee for emphasizing the need for a heteroscedastic-consistent test. Clearly, the coefficient estimates from different cross-sections are unlikely to have the same variance. Hence, except in the case of testing zero means for the coefficient vectors, the routine and widespread reliance on conventional t-tests and F-tests is inappropriate.

Table II

Persistence Estimates: Robustness to Sample Period and Seasonality

Construction of estimates reported in this table follows Table I. For parsimony, we only report the results with the "residual" performance proxied by the residual from the market model regressions using the equally weighted NYSE index; the results with other benchmarks are similar.

	Time-Average of Cross-Sectional Coefficients (t-Statistics in Parentheses)									
	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{a}_4$	$\hat{a}_5$	$\hat{a}_6$	$\hat{a}_7$	$\hat{a}_8$	$\Sigma\hat{a}_i$	χ^2-Test[a]
1974–1988	0.05	0.10	0.06	0.11	−0.05	−0.05	−0.06	0.04	0.18	
for comparison	(1.03)	(2.56)	(1.62)	(2.88)	(−1.69)	(−1.84)	(−1.79)	(1.61)		
Panel A. Sample Halves										
1974–1981	0.02	0.13	0.09	0.08	−0.06	−0.04	−0.03	0.05	0.24	H_0: equal
	(0.43)	(2.46)	(2.38)	(1.58)	(−1.30)	(−1.10)	(−0.71)	(1.43)		across
1981–1988	0.07	0.06	0.02	0.13	−0.04	−0.07	−0.09	0.03	0.12	split-sample
	(0.95)	(1.19)	(0.34)	(2.46)	(−1.07)	(−1.47)	(−1.72)	(0.82)		5.11
										[0.75]
Panel B. Calendar Quarter Subsamples										
Quarter 1	0.12	0.14	0.13	0.01	−0.04	−0.15	−0.21	0.13	0.13	
	(1.13)	(1.67)	(2.79)	(0.16)	(−0.60)	(−2.49)	(−2.91)	(3.38)		
Quarter 2	0.14	0.14	0.11	0.06	−0.12	−0.01	0.04	0.02	0.37	H_0: equal
	(2.24)	(3.31)	(2.18)	(0.90)	(−1.82)	(−0.24)	(0.66)	(0.42)		across
Quarter 3	−0.16	0.08	0.04	0.18	−0.05	0.00	−0.03	0.05	0.12	quarters
	(−2.06)	(1.03)	(0.35)	(2.20)	(−1.01)	(0.08)	(−0.52)	(0.87)		107.89
Quarter 4	0.08	0.02	−0.06	0.17	0.01	−0.04	−0.03	−0.04	0.10	[0.00]
	(0.93)	(0.27)	(−2.12)	(3.40)	(0.20)	(−0.89)	(−0.56)	(−1.21)		

[a]The χ^2-statistic tests the joint hypothesis that each of the slope coefficients (lags 1 to 8) is equal between the subsamples in the panel. The subsamples to which the χ^2-test applies is indicated immediately above the statistic (see H_0). The p-value is reported in brackets.

quarter) seasonality since, in results not reported, we reject the hypothesis of equal coefficients across subsamples of quarters two, three, and four.

In sum, the evidence from Tables I and II indicates statistically significant positive persistence between performance from one quarter to the next. The persistence coefficients are significantly positive for four quarters, although different across quarters. The persistence fades away beyond a year, which is consistent with a hot (and icy) hands phenomenon.

II. Performance Potential of Strategies That Exploit Short-Term Persistence

We evaluate the performance, during 1975–88, of executable strategies that exploit hot hands. The outcomes from these strategy simulations provide an economic assessment of the statistical persistence finding.

For every quarter in our sample period, we distribute the eligible sample of funds into eight performance-ranked portfolios (about 12 to 15 mutual funds per portfolio).[11] The first-octile portfolio is composed of the poorest performers in the recent evaluation period, the second-octile portfolio is composed of the next-best performers, and so on. We use the net return over an evaluation period for fund ranking, eschewing potential exploitation of the α-estimates of equation (2) in favor of simplicity. We refer to portfolios so constructed as rank portfolios.

In the terminology of Treynor and Black (1972), we start by evaluating the *active* rank portfolios with respect to the *passive* benchmarks. We consider two evaluation criteria and different benchmark portfolios:

1. Jensen's α (excess expected return controlling for premiums arising for covariance risks measured with the benchmarks). We compute the following regression by OLS:

$$r_{pt} = \alpha_p + \sum_{k=1}^{K} \beta_{kp} B_{kt} + \varepsilon_{it}. \tag{4}$$

Here r_{pt} is the excess return on portfolio p in quarter t; p is either a hot-hands octile portfolio, or a zero-investment best-minus-worst portfolio. The zero investment strategy that is long in the top performers' portfolio and short in the worst performers' portfolio, denoted the "maximal" portfolio, measures the maximal gain from exploiting performance

[11] The exact allocation procedure is as follows. In quarter t, order the N_t available funds based on their net returns in the quarter. Let the rank of fund i's return be $rank(r_i)$. The fund is assigned to octile j such that the following formula is satisfied:

$$(j-1)\lfloor N_t/8 \rfloor + \sum_{k=1}^{j-1} \delta_k \le rank(r_i) < j\lfloor N_t/8 \rfloor + \sum_{k=1}^{j} \delta_k, \text{ where } \delta_k = \begin{cases} 1 \text{ if } k \le N_t \bmod 8, \text{ and} \\ 0 \text{ otherwise.} \end{cases}$$

Here $\lfloor \cdot \rfloor$ denotes the integer portion of the enclosed fraction. The number of funds in octile k in quarter t is given by $\delta_k + \lfloor N_t/8 \rfloor$.

persistence. The intercept from regression (4), α_p, provides an estimate of Jensen's α.[12] The B_{kt}'s are the excess returns on the benchmark portfolios, and β_{kp}'s are the sensitivities (betas) of the returns of portfolio p to the benchmark returns; generally $K = 1$, but $K = 8$ for the multi-portfolio benchmark, P8.

We evaluate the significance of the α estimates by a heteroscedastic-consistent z-statistic following White (1980). (The rank portfolios are heteroscedastic because the funds included in each portfolio vary both in number and in identity from quarter to quarter. More generally, Breen *et al.* (1986) have shown the need for heteroscedasticity corrections in performance evaluation studies, and report that White's correction performs well.)

2. Spearman's statistic (a nonparametric test of the predictability of performance ranks). We compute Spearman's statistic using the sum of the squared differences for each portfolio's octile rank from its α-performance rank in the next period. If the hot hands strategy can perfectly sort funds ex ante into performance octiles, Spearman's statistic will be zero; if it has no ex ante sorting capability, the expected value of Spearman's statistic will be 84 ($= (N^3 - N)/6$ for $N = 8$).[13]

A. Econometric Results on the Performance of Rank Portfolios

Consider rank portfolios with evaluation periods ranging from one quarter to eight quarters. (For example, a four-quarter evaluation period indicates that the funds have been assigned to hot hands portfolio octiles on the basis of their relative returns in the most recent four quarters.) If the evaluation period is too short, we expect the signal of superior performance due to skill will get lost in the noise from chance factors. If the evaluation period is too long, the salience of hot hands will diminish.

In Table III, we report on summary statistics, Sharpe's measure, Jensen's α, and Spearman's statistic. The following features are common at each of the four evaluation periods:

1. The mean excess return increases monotonically with the octile ranks. A portfolio of better (worse) recent performers does better (worse) in the next quarter than the mean fund performance.
2. Sharpe's measure, the ratio of mean excess return to standard deviation (total risk), also increases monotonically with the octile ranks.

[12] If managers engage in market-timing strategies (which lead to time-varying betas of their portfolios), then the intercept from a simple regression like (4) cannot provide appropriate inference as to their stock selection ability. Even if managers do not engage in market timing, the estimation of Jensen's measure of stock selection skills is complicated because fund managers will change their portfolio holdings in response to the arrival of private information and will lead to time-varying betas (Admati and Ross (1985), and references therein). In practice, fortunately, the theoretical possibilities of time-varying betas seem not to matter in the case of mutual funds (Grinblatt and Titman (1989a, 1989b)).

[13] The p-values for Spearman's statistic are obtained from Lehmann (1975).

3. Jensen's α rises monotonically with octile rank, independent of the benchmark considered. The strong performance-sorting capacity of the hot hands strategy is clearly seen with the Spearman's statistic: it is always extremely significant (p-values below 1%), independent of the benchmark used.
4. The estimates of Jensen's α are similar with the VWCRSP, EWMF, or P8 benchmarks: The α-estimates are positive for the top performers' octiles and negative for the poor performers' octiles.
5. The benchmark choices systematically affect the evaluation of mutual funds' portfolios as follows: (a) The betas are about unity with VWCRSP (or EWMF); in results not reported, the beta estimates of the octile portfolios are substantially lower (about 30%) with EWNYSE than with VWCRSP. (b) Jensen's α is lower by about 40 to 60 basis points when the benchmark is EWNYSE rather than VWCRSP (or EWMF or P8).

Patterns that vary by evaluation periods are as follows:

1. Systematic risk (beta) is the same across octile portfolios for the one- and two-quarter evaluation periods. However, it increases about 3 to 4% per octile rank for the four- and eight-quarter evaluation periods. This pattern is clearly evident in the betas of the zero-investment maximal portfolio: its beta is insignificantly different from zero for the shorter evaluation periods, but clearly positive for the longer evaluation periods.
2. Jensen's α for the maximal portfolio, always positive, becomes statistically significant for the two-, four- and eight-quarter evaluation periods (with a point estimate of 6 to 8% per year). The significance is weaker with the P8 benchmark, which might be due to the smaller sample period (only 70% of that otherwise available) and the loss of degrees of freedom (since seven additional coefficients need to be estimated).
3. The impact of the hot hands strategies reaches a maximum with a four-quarter evaluation period, as evidenced by the magnitudes and relative significances of the Jensen's α's. This finding is consistent with the results of Table I, where there was significant positive persistence up to four lags in the residual returns. At a four-quarter evaluation period, the poor performers' octiles have significantly negative α's, though the best performers' octiles do not obtain statistically significant positive α's.

Overall, Table III establishes that the hot hands strategy can clearly sort funds into relative performance ranks. It definitely identifies underperformers. The point estimates of α's for the eighth octile with a four-quarter evaluation indicate that an active portfolio of mutual funds can outperform all the benchmarks. The superior performance of this portfolio is statistically significant only in relation to fellow funds, which is the relevant benchmark for most mutual fund investors.

Since the composition of the rank portfolios changes from quarter to quarter, the assumed constancy of their betas over the sample period may be

of concern. We assess this issue by computing an indirect time-varying beta for each octile portfolio: in each quarter, the beta of the portfolio is set equal to the mean of the beta estimates of the included funds. This approach assumes that the betas of the mutual funds remain approximately constant over our sample period. In results not reported, for every octile portfolio, we find that the indirect beta is always very similar to the directly estimated time-invariant beta from the market model regression (4). Moreover, the inferences of performance persistence do not change—indeed, they strengthen slightly—if we correct for beta-risk using the indirect measure. We establish further the robustness of the persistence results in the following subsections.

B. Simulations: Exploring Selectivity, Timing, and Sample Artifacts

The statistical tests in Section I favor the existence of short-term persistence (Hypothesis 3). Some of the findings in Table III, however, could be driven by permanent persistence (Hypothesis 2). To assess the economic size of the short-term persistence component alone, we perform several sets of additional tests and simulations. These explorations are designed to distinguish between pure selectivity (i.e., sorting into good versus bad funds unconditional on recent short-term performance) and timing selectivity.

Consider the top octile portfolio with a four-quarter evaluation period. To what extent is its performance in our sample period due merely to picking superior performing funds? To address this question, consider the results from 5000 simulations.

Each simulation generates quarterly returns on eight octile portfolios spanning the same sample period in Table III. For each simulation, a fund's assignment to an octile in any quarter is stochastic, with the probability of being assigned to an octile set equal to the relative frequency that the fund actually fell into the octile in Table III (cumulated over the entire sample period). For example, consider the case of the well-known 44 Wall Street fund. In the course of constructing the octile returns presented in Table III, it got assigned to the bottom (top) octile in 45% (38%) of the quarters. For the simulations, the return of 44 Wall Street fund in any quarter would be assigned to the bottom (top) octile with probability 45% (38%) of the time. Thus, a fund that frequently wound up in the bottom (top) octile for the hot hands assessment in Table III would do so as well in the simulations; however, the particular quarters in which it got assigned to the bottom (top) octile would vary randomly between simulations.

By comparing the performance measures for the octile portfolios shown in Table III with those generated in the simulation, we can assess whether the timing of inclusion in the different octiles (i.e., short-term performance persistence) is an important factor for the results. If timing (picking the right fund at the right time, rather than simply picking overall superior funds) is an important feature underlying our results, the performance statistics in Table III will lie in the tails of the simulated distributions. Table IV reports the fractions of the simulation distributions that lie below selected statistics

Table III

Rank Portfolios: Summary Statistics, Jensen's α, and Sharpe's Measure

Eight rank portfolios, reconstituted each quarter, are formed based on performance ranks for evaluation periods of one, two, four, and eight quarters. The overall sample period is 1974Q1–1988Q4. Summary statistics for each of the octile portfolios and a best-worst portfolio (which is long in the eighth octile (best performers in evaluation period) and short in the first octile (worst performers)) are reported below. For each of four benchmarks, Jensen's α measure of performance for each rank portfolio is shown. The Spearman's statistic in the last column provides a measure of the predictability in the rank ordering of α's of the octile portfolios. Spearman's statistic is computed as the sum of the squared differences between the octile's construction rank and its overall post-construction α-rank.

	Results on Rank Portfolios[a]									
	1 (Worst)	2	3	4	5	6	7	8 (Best)	Best-Worst Portfolio[b]	Spearman's Statistic[b]
Panel A. One-Quarter Evaluation Period										
Mean excess return	1.42	1.66	1.84	1.91	2.11	2.45	2.41	2.67	1.25	
β (VWCRSP)	1.01	0.99	0.98	0.99	0.98	1.05	1.06	1.09	0.08	
Sharpe's measure	0.14	0.18	0.21	0.21	0.24	0.26	0.24	0.25	0.19	
α (EWNYSE)	−1.59	−1.17	−0.90	−0.86	−0.62	−0.53	−0.53	−0.42	1.17	2
	(−2.72)	(−3.15)	(−2.55)	(−2.35)	(−1.70)	(−1.47)	(−1.12)	(−0.69)	[0.17]	[0.00]
α (VWCRSP)	−0.81	−0.52	−0.31	−0.27	−0.05	0.13	0.08	0.26	1.07	2
	(−1.36)	(−1.57)	(−1.34)	(−1.29)	(−0.26)	(0.48)	(0.21)	(0.44)	[0.19]	[0.00]
α (EWMF)	−0.67	−0.36	−0.13	−0.09	0.13	0.30	0.23	0.37	1.03	2
	(−1.26)	(−1.31)	(−0.70)	(−0.55)	(0.71)	(1.71)	(0.85)	(0.83)	[0.22]	[0.00]
α (P8)[c]	−0.68	−1.19	−0.54	−0.32	−0.39	−0.21	0.44	0.95	1.63	4
	(−1.11)	(−2.55)	(−1.75)	(−1.24)	(−0.97)	−0.56)	(0.85)	(0.99)	[0.22]	[0.00]
Panel B. Two-Quarter Evaluation Period										
Mean excess return	1.33	1.73	1.81	1.99	2.07	1.99	2.33	3.24	1.91	
β (VWCRSP)	1.04	1.01	1.00	0.98	1.01	1.03	0.97	1.10	0.07	
Sharpe's measure	0.13	0.19	0.20	0.22	0.23	0.21	0.26	0.29	0.26	
α (EWNYSE)	−1.81	−1.16	−1.03	−0.77	−0.79	−0.86	−0.39	0.20	2.01	8
	(−3.39)	(−3.18)	(−2.90)	(−2.18)	(−2.24)	(−2.09)	(−0.87)	(0.32)	[0.02]	[0.00]
α (VWCRSP)	−0.96	−0.49	−0.40	−0.17	−0.16	−0.28	0.18	0.80	1.76	6
	(−1.68)	(−1.79)	(−1.64)	(−0.78)	(−0.59)	(−1.06)	(0.49)	(1.32)	[0.04]	[0.00]
α (EWMF)	−0.80	−0.31	−0.22	0.00	0.01	−0.11	0.31	0.90	1.70	6
	(−1.54)	(−1.27)	(−1.17)	(0.01)	(0.07)	(−0.62)	(1.30)	(2.01)	[0.04]	[0.00]
α (P8)[c]	−0.83	−0.89	−0.28	−0.46	−0.65	−0.36	0.43	1.09	1.92	16
	(−1.22)	(−2.62)	(−0.73)	(−1.06)	(−2.34)	(−0.84)	(0.84)	(1.19)	[0.18]	[0.00]

Table III—*Continued*

	Results on Rank Portfolios[a]									
	1 (Worst)	2	3	4	5	6	7	8 (Best)	Best-Worst Portfolio[b]	Spearman's Statistic[b]
Panel C. Four-Quarter Evaluation Period										
Mean excess return	0.99	1.68	1.40	1.82	1.90	2.32	2.94	3.47	2.48	
β (VWCRSP)	0.93	0.98	0.97	0.99	1.00	1.00	1.08	1.18	0.25	
Sharpe's measure	0.10	0.19	0.16	0.20	0.21	0.25	0.29	0.30	0.36	
α (EWNYSE)	−1.85	−1.09	−1.33	−0.94	−0.89	−0.51	−0.10	0.14	1.99	2
	(−3.54)	(−3.06)	(−3.63)	(−2.57)	(−2.35)	(−1.26)	(−0.24)	(0.24)	[0.01]	[0.00]
α (VWCRSP)	−1.07	−0.49	−0.74	−0.37	−0.31	0.10	0.56	0.86	1.93	2
	(−1.91)	(−2.12)	(−3.19)	(−1.85)	(−1.21)	(0.33)	(1.43)	(1.49)	[0.01]	[0.00]
α (EWMF)	−0.92	−0.31	−0.56	−0.19	−0.13	0.25	0.70	0.98	1.90	2
	(−1.78)	(−1.48)	(−2.88)	(−1.15)	(−0.72)	(1.36)	(2.96)	(2.56)	[0.01]	[0.00]
α (P8)[c]	−0.75	−0.80	−0.89	−0.51	−0.40	−0.01	0.68	0.71	1.45	8
	(−1.15)	(−2.16)	(−2.81)	(−1.64)	(−1.83)	(−0.03)	(1.12)	(0.93)	[0.21]	[0.00]
Panel D. Eight-Quarter Evaluation Period										
Mean excess return	0.69	1.05	1.40	1.76	1.59	2.24	2.11	2.80	2.11	
β (VWCRSP)	0.88	0.91	1.00	1.05	1.00	1.08	1.10	1.16	0.28	
Sharpe's measure	0.09	0.14	0.17	0.20	0.19	0.24	0.22	0.26	0.35	
α (EWNYSE)	−1.49	−1.18	−1.03	−0.80	−0.87	−0.46	−0.60	−0.15	1.35	4
	(−3.38)	(−3.35)	(−2.98)	(−2.30)	(−2.52)	(−1.37)	(−1.43)	(−0.26)	[0.05]	[0.00]
α (VWCRSP)	−0.89	−0.57	−0.37	−0.11	−0.20	0.32	0.15	0.74	1.62	4
	(−1.90)	(−2.58)	(−1.84)	(−0.42)	(−0.82)	(0.97)	(0.41)	(1.20)	[0.02]	[0.00]
α (EWMF)	−0.80	−0.46	−0.25	0.01	−0.09	0.41	0.25	0.77	1.57	4
	(−1.98)	(−2.07)	(−1.36)	(0.07)	(−0.49)	(2.34)	(1.11)	(1.91)	[0.02]	[0.00]
α (P8)[c]	−0.66	−0.12	−0.16	−0.14	−0.04	0.36	0.29	1.28	1.95	8
	(−1.46)	(−0.42)	(−0.40)	(−0.56)	(−0.11)	(0.90)	(0.75)	(1.62)	[0.07]	[0.00]

[a]White's z-statistic, which corrects for the heteroscedasticity of the rank portfolio returns, is reported in parentheses below the estimates of Jensen's α.

[b]The p-values of observing the estimated statistic under the null hypothesis of no predictability is reported in brackets.

[c]The P8 benchmark is available only for 1975–1984, and therefore results with this benchmark are based on subperiods ending in 1984Q4.

Table IV

Timing Versus Selectivity for the "Four-Quarter Evaluation" Strategy

This table reports the percentage of values from 5000 simulations that lie below the estimated statistic shown in Panel C of Table III. The simulations assign funds to each octile in each quarter based on the overall frequency with which they were assigned to that octile (rather than on recent fund performance as in the hot hands strategy). If the observed statistics in Table III lie in the middle of the corresponding simulated distributions, then the hot hands results can be attributed to selectivity (that is, picking the right fund). If, however, the observed values lie in the tails of the simulated distributions, then the findings are more likely due to timing (picking the right fund at the right time).

	α-Estimate for Rank Portfolios			Spearman's
Benchmark	Octile 1(%)	Octile 8(%)	Best-Worst(%)	Statistic
EWNYSE	19	96	94	3
VWCRSP	38	93	86	6
EWMF	41	92	85	6
P8[a]	71	85	61	7

[a]The P8 benchmark is available only for 1975–1984, and therefore simulations for this benchmark cover only subperiods ending in 1984Q4.

from Panel C of Table III. We find, for example, that the estimated α's for the top octile and the best-worst portfolios are in the upper tail of the simulated distributions. The Spearman statistics, assessing rank orders across octiles, fall in the lower tail. These results are consistent with an economically important timing component for achieving superior performance, which corroborates the statistical persistence identified in Table I.

In contrast, the significant underperformance of the bottom octile portfolio reported in Table III appears to be driven by sustained poor-performing funds. Table IV reports that the α-estimates for the underperforming bottom octile portfolio are in the center of the simulated distributions, which control for performance over the entire sample period. This suggests that pure fund selection is important in understanding the observed relative and absolute inferior performance of the bottom octile. (Sustained inferior performers are easily explained if some funds without superior skills churn their portfolios too much and incur relatively high expenses which lowers their net performance rank over the sample period.)

To address further the possibility that our results may be driven mainly by sustained underperformers in our sample, in results not presented, we analyzed a subsample that removed the forty worst-performing mutual funds (bottom quartile) over the sample period.[14] We replicated the statistics reported in Table III, Panel C, and continued to identify performance persistence. For example, the α-estimate for the best-worst portfolio fell only by 20

[14]The criterion we use is mean excess returns. Removal on the basis of lowest alphas produces nearly identical results.

basis points against each of our single-portfolio benchmarks, which remained significant at the 5% level as did the Spearman statistic. (There is no need to replicate Table I with such a subsample since the patterns in Table I were established with deviations from the sample mean of each fund's own performance level.)

In unreported results, against the P8 benchmark there was a larger erosion of the performance differential (40 basis points) between top and bottom. But recall that in Table IV we found little to no timing component for the underperformance of the lowest octile portfolio versus the P8 benchmark. Thus, the removal of underperforming funds can be expected to reduce substantially the performance differential between the best and the worst funds in the remaining subsample. Nonetheless, there continued to be substantial sorting of performance: the Spearman statistic with the P8 benchmark for this subsample proved significant at the 5% level.

In other results not shown, we evaluated whether the hot hands findings could have arisen through a spurious interaction of our hot hands strategy with the time-series properties of equity returns during 1974–87. We generated 100 artificial portfolios, each of which was an equally weighted portfolio of 100 equities drawn randomly from those listed on the monthly returns tapes distributed by the Center for Research in Security Prices (CRSP), University of Chicago. For this set of 100 unmanaged portfolios over 1974–87, we simulated the best-performers' octile portfolio strategy with a four-quarter evaluation period. One hundred such simulations were carried out. With the unmanaged portfolios, we found that the octile 8 equivalents obtain a virtually zero Jensen's α using the VWCRSP or the EWMF benchmarks. We conclude that the hot hands finding is unlikely to be an artifact of the behavior of equities in the particular sample period.[15]

Overall, short-run persistence (timing component) in our sample of fund returns delivers superior performance when compared to all single portfolio benchmarks, though pure selectivity is important for underperformers. Using single portfolio benchmarks (such as the EWNYSE), the timing component can deliver a best-worst relative performance of about 6%. Pure timing in the selection of mutual funds does not offer superior performance relative to the P8 benchmark, which is constrained to a smaller time period and requires estimating more nuisance parameters, though timing still delivers relative ordering among funds.

C. Nonlinearities and Option Characteristics

So far, we have implicitly assumed that the equilibrium-generating process for fund returns is linearly related to benchmark returns. If some mutual fund managers engage in market timing, the returns of their funds will have

[15] In a related exploration, we studied the subperiod of 1974:1–1987:3 in order to exclude the October '87 market crash and subsequent developments. There is little change. If anything, we notice a slight improvement in the size and significance of results corresponding to Table III with the pre-crash subperiod. Results shown earlier in Tables I and II are similarly unaffected.

option-like nonlinear characteristic lines—see Treynor and Mazuy (1966), Henriksson and Merton (1981), or more recently Connor and Korajczyk (1991). We report below some evidence of nonlinear option-like features displayed by our rank portfolios. Correcting for it, however, leaves our main results unaffected.

Consider the following extension to regression (4) that is motivated by Henriksson and Merton (1981):[16]

$$r_{pt} = \alpha_p + \beta_p B_t + \gamma_p \text{put}(B_t) + \varepsilon_{pt}, \tag{5}$$

where $\text{put}(B_t) = \max(-B_t, 0)$. The γ can be nonzero if (i) the mutual fund displays market-timing ability, or (ii) the mutual fund includes options (most likely to be synthetically created by dynamic trading) on the benchmark B_t, or (iii) the equilibrium model of asset pricing is nonlinear, or (iv) risky debt is present, which induces put-like features in equities. In the discussion that follows, we ignore the case of a nonlinear equilibrium asset-pricing relation or the consequences of bankruptcy risk, leaving that for future research.[17]

In Panel A of Table V, the γ-estimates with the different single benchmarks are reported for the different rank portfolios. With the EWNYSE benchmark, the γ-estimate is positive but not significantly different from zero for the first octile. It becomes negative for the higher octiles (and significantly so for some of them), which is consistent with the higher octile portfolios having implicitly reduced upside potential (i.e., implicitly having sold a put on the benchmark or having sold "portfolio insurance"). However, with the VWCRSP benchmark, the nonlinearity (as evidenced by significant γ) is not seen (or with the EWMF, not reported for brevity). This is consistent with our earlier suggestion that mutual fund evaluation with the EWNYSE benchmark may be problematic.[18]

[16] We do not assess nonlinearities with the P8 benchmark (which has 8 portfolios) since we would have to estimate 16 coefficients in the following regressions, and have available a smaller sample period (because P8 is not available after 1984).

[17] A three-moment asset-pricing model is developed by Kraus and Litzenberger (1976). Lim (1989) finds suggestive evidence to support it, at least in the long run. Jagannathan and Korajczyk (1986) discuss the put-like properties of equities in the presence of risky debt, and some consequences for performance measurement.

[18] The negative γ's for the top performers' octiles could explain their superior performance measured within a linear model framework, relative to the poor performers' octiles. Suppose that the hot hands strategy simply sorts among funds on the basis of those whose positions are relatively short in portfolio insurance. Now, even if the benchmark(s) are mean-variance efficient with respect to equities, a linear factor model with zero intercept adequately describes excess equity returns and market prices are efficient so that (implicit) portfolio insurance is fairly priced, the estimation of regression (4) will exhibit a positive intercept. In fact, if fund managers know that they are evaluated on Jensen's α as computed using regression (4), they can game against the evaluation criteria by adopting a dynamic trading strategy that implicitly makes them short on a portfolio insurance position—for instance, see Dybvig and Ingersoll (1982). Similarly, if fund managers are naively evaluated by counting the number of times they outperform the benchmark, then they can game against the evaluation method by adopting option-like positions.

Table V

Correction for Nonlinear Option-Like Characteristics

Four-quarter evaluation period; the overall sample period is 1974Q1–1988Q4. Rank portfolios are formed as in Table III. In Panel A, we report results derived by regressing the returns of the rank portfolios on the returns of a benchmark portfolio and the returns from holding a put option (with a strike price of zero and quarter-end expiration) on the benchmark portfolio—see regression (5) in the text. The γ-estimate is the coefficient of the put returns; it indicates the degree to which the rank portfolios exhibit nonlinearities in their characteristic lines, possibly due to market-timing ability. In Panel B, we consider the result from a similar regression—see regression (6) in the text—where the return on the put position is computed *net* of the ex ante cost assigned using the Black-Scholes formula. This approach provides a performance measure, α-estimate, that accounts for a nonlinearity in the characteristic line.

Benchmark Choice[a]	Results on Rank Portfolios[b]								Best-Worst Portfolio[c]
	1	2	3	4	5	6	7	8	
Panel A. Extent of Nonlinearity in Characteristic Line: γ-Estimates (Regression (5))									
EWNYSE	0.241	−0.193	−0.167	−0.320	−0.349	−0.439	−0.434	−0.412	−0.65
	(1.43)	(−1.54)	(−1.43)	(−1.83)	(−2.23)	(−2.60)	(−1.91)	(−1.59)	(−1.75)
VWCRSP	0.498	0.079	0.108	−0.039	−0.035	−0.161	−0.082	0.051	−0.45
	(1.67)	(1.17)	(1.29)	(−0.37)	(−0.30)	(−1.37)	(−0.48)	(0.23)	(−1.95)
Panel B. α-Estimates After Netput Corrections (Regression (6))									
EWNYSE	−1.844	−1.095	−1.337	−0.954	−0.904	−0.526	−0.121	0.125	1.97
	(−3.69)	(−3.05)	(−3.61)	(−2.73)	(−2.44)	(−1.48)	(−0.29)	(0.22)	(2.68)
VWCRSP	−1.341	−0.528	−0.796	−0.348	−0.290	0.191	0.601	0.836	2.18
	(−2.66)	(−2.32)	(−3.57)	(−1.78)	(−1.18)	(0.62)	(1.53)	(1.39)	(2.90)

[a] The results with EWCRSP are very similar to those shown with the EWNYSE, while the results with EWMF are very similar to those shown with VWCRSP.

[b] White's z-statistics, which are heteroscedastic-consistent, are shown in parentheses below the coefficient estimates.

[c] The best-worst portfolio arises from a zero net investment strategy of going short the first octile (worst performers) and investing the proceeds in the eighth octile (best performers).

Following Connor and Korajczyk (1991), a simple correction is made to account for the nonzero γ's that we find. We construct a variable, netput, which is the payoff to a European put on the benchmark (which expires at the end of the quarter, and has a strike price equal to the beginning-of-quarter level) minus the Treasury bill return necessary to pay the price of the put. We use the Black-Scholes formula to price the put option on the benchmark; for the formula, the required inputs of the risk-free rate and the benchmark's variance are taken as the Treasury bill yield and the benchmark's sample variance over the 1975–88 period respectively. The performance evaluation regression is:

$$r_{pt} = \alpha_p + \beta_p B_t + \gamma_p \text{netput}(B_t) + e_{pt}. \tag{6}$$

The intercept, α_p, can be interpreted as the net measure of superior performance after correcting for a fair market valuation of any (implicit) options in the portfolio.

The estimates of α_p after the netput adjustment are found in Panel B of Table V. The maximal gains from following the hot hands strategy are unchanged relative to the results in Panel C of Table III. The relation in Table III between octile ranks and α's continues to hold clearly: Spearman's statistic is a very low 2, which has a p-value below 1% for all the benchmarks.[19]

III. Other Samples and Survivorship-Biased Subsamples

The reinforcement of the results across Sections I and II is less than additive since the results are drawn from the same data set. Thus we examine a different sample in subsection III.*A* to confirm our findings. In subsection III.*B* we explore the sensitivity of persistence inferences to subsamples purposely selected to induce survivorship bias.

A. Other Samples

A different sample from our own, based on 130 equity mutual funds over 1968–82, is used by Henriksson (1984), Lehmann and Modest (1987), and Connor and Korajczyk (1991)—we refer to it hereafter as the HLM sample. Unlike our sample, the funds in the HLM sample have a mix of objectives, and are survivors over the sample period. Further, some of the included funds may have changed their strategy and/or adopted loads and/or imposed investment restrictions over subperiods. Keeping these caveats in mind, we examine the hot hands strategy utilizing a four-quarter evaluation using this sample.

[19] In other results not reported, the β's that are estimated in regressions (5) and (6) no longer increase monotonically with octile rank as they do in Table III.

Table VI

Results on Rank Portfolios Employing the Henriksson-Lehmann-Modest Sample

Four-quarter evaluation period: the data cover 1968–1982. This table parallels Table III, but uses a different sample of mutual funds that was constructed by Henriksson (1984) and updated by Lehmann and Modest (1987). Eight rank portfolios, reconstituted quarterly, are formed using a four-quarter evaluation period. Octile 1 has the worst performers while octile 8 has the best performers.

	Results on Rank Portfolios[a]								Best-Worst
	1	2	3	4	5	6	7	8	Portfolio[a]
Mean excess return	−0.62	0.46	−0.10	−0.15	0.32	0.33	0.93	1.00	1.62
Sharpe's measure	−0.06	−0.05	−0.01	−0.02	0.04	0.04	0.11	0.11	0.27
β (VWCRSP)	1.06	0.97	0.94	0.92	0.93	0.87	0.86	0.88	−0.17
α (EWCRSP)	−1.54	−1.29	−0.88	−0.92	−0.47	−0.42	0.19	0.21	1.76
	(−2.21)	(−2.25)	(−1.48)	(−1.59)	(−0.86)	(−0.81)	(0.35)	(0.36)	(2.35)
α (VWCRSP)	−0.88	−0.69	−0.33	−0.38	0.09	0.11	0.72	0.79	1.67
	(−2.04)	(−3.01)	(−1.58)	(−1.98)	(0.49)	(0.50)	(2.37)	(1.71)	(2.22)
α (EWMF2)[b]	−0.79	−0.61	−0.24	−0.30	0.17	0.19	0.80	0.86	1.65
	(−2.03)	(−2.98)	(−1.73)	(−2.32)	(1.69)	(1.32)	(3.38)	(2.10)	(2.19)

[a] White's z-statistics, which are heteroscedastic-consistent, are shown in parentheses below the α-estimates. Asymptotically, the z-statistics have a standard normal distribution.

[b] The EWMF2 benchmark is the average fund return in the Henriksson-Lehmann-Modest sample.

The results in Table VI are remarkably similar to the earlier results in Table III despite quite different samples. The relation between octile rank and performance is near monotonic and as strong as before—Spearman's statistic, not reported, is 2 with all of the benchmarks, which has a p-value of less than 1%. The best-worst portfolio has a positive and significant Jensen's α of around 1.7% per quarter, which is in the same ballpark as the earlier estimate of around 1.9%. As before, the Jensen's α's for the low octile portfolios are individually significant while those for the top octile portfolios, though positive, are generally not significant.

Finally, consider a bit of out-of-sample evidence based on the funds in our sample during 1989–90. The average mutual fund performance (EWMF) for 1989–90 was 2.24% per quarter. For this admittedly small period, we simply divide the funds into above-median performers and below median performers based on the funds' performances in the previous four quarters. We update the selection every quarter. The average return per quarter of this simple hot hands strategy for the eight quarters during 1989–90 period is as follows:

1. superior performers (above median in previous four quarters): 2.46%;
2. inferior performers (below median in previous four quarters): 2.01%.

The hot hands strategy discriminates ex ante among winners and losers during 1989–90.

B. Survivorship-Biased Subsamples

In empirical studies, any interaction between the idiosyncrasies of the sample construction process and the sample's properties is of concern. For fund performance evaluation, one concern relates to samples that only include funds which survive till the end of the study period (which we denote as survivors). As Brown *et al.* (1991, p. 7) note straightforwardly regarding survivors' samples: "A manager who takes on a great deal of risk will have a high probability of failure. However, *if* he or she survives, the probability is that this manager took a large bet and won. High returns persist... this is a total risk effect; risk adjustment using beta or other measure of non-idiosyncratic risk may not fully correct for it." Brown *et al.* use simulated data to demonstrate spurious persistence measurement is induced in survivors' samples where (1) a beta-type correction does not fully account for the different risks of funds, and (2) funds that substantially underperform in a simulation period cease to exist.

However, as Brown *et al.* recognize, the simulation results overstate the spurious persistence in a survivors' sample if, as is likely in practice, performance must be inferior for a sustained period before a fund ceases to exist.[20] Also, the survivorship bias will be small if few funds terminate due to sustained poor performance; witness the long survival of the significantly underperforming 44 Wall Street fund. Using our sample, we shed some light on the practical magnitude of bias in persistence measurement, if any, from a survivorship filter.

The persistence from our entire sample, which is constructed to minimize survivorship bias, provides a benchmark. For quarter t, let π_{it} be the percentile performance rank of fund i relative to the other funds in our sample. Given a subsample S, we compute two types of average percentile rank for funds as follows:

$$\pi_S^j = \sum_{i \in S} \sum_{t \in T_i^j} \pi_{it};\ j = 1, 2. \tag{7}$$

For $j = 1$, T_i^1 includes all the quarters for which valid data on fund i is available; and for $j = 2$, T_i^2 includes the four quarters prior to and including the last quarter of data available for fund i. We report π_S^1 and π_S^2 for different subsamples S in Panel A of Table VII. In addition to survivor's subsamples, we show results from the subsample of funds that adopt a sales load.

Funds that were terminated do quite a bit worse than average ($\pi_S^1 = 23$), although their performance in the four quarters immediately preceding their

[20] Patel, Zeckhauser, and Hendricks (1992) document that poor fund performance leads to outflows of money from the fund, although not in the magnitudes required to close down a fund as modeled by Brown *et al.* (1992).

cessation is relatively better ($\pi_S^2 = 37$).[21] In contrast, funds that switch to adopting load charges do slightly better than average overall ($\pi_S^1 = 52$), and somewhat better in their four quarters prior to their switch ($\pi_S^2 = 56$). Overall, funds whose return history in our sample ends before 1988 for any reason (and would have been dropped in a survivors' sample) have a π_S^1 of 41 and π_S^2 of 45, results that are little worse than the π_S^1 of 51 for the survivors.

Some further results related to "survivorship bias" follow. In Panel B of Table VII, we compare the persistence coefficient estimates that we had computed in Table I with those that would have been obtained using survivors' subsamples. (For brevity, only results with the residuals relative to the EWNYSE proxy are shown. Results with other benchmarks are similar.) We consider a subsample of the 51 funds that have complete data for our sample period, and a more inclusive subsample of the 94 funds that remain by the end of 1988 (regardless of the date of their starting period in the sample). The magnitude and patterns of performance persistence for these subsamples are remarkably similar to those based on the full sample. The hypothesis of the equality of persistence coefficients across the three samples is not rejected for any lag.

We also provide some results from bootstrap simulations that verify the significance of the persistence coefficients. Proceeding under the null hypothesis of no serial dependence in fund performance, we assume that the cross-section of excess fund returns are independent draws from a common but unknown distribution Φ. Of interest is the distribution of the mean persistence coefficients, which we computed in Table I and Panel B of Table VII, under the null hypothesis of zero values, i.e., $\Psi(\hat{a} - 0) \equiv \Psi(\hat{a})$. Here Ψ depends on Φ, and other factors which we treat as fixed (such as the number of persistence coefficients estimated, the number of cross-sections, and the number of funds per cross-section). Our implementation of the bootstrap method uses resampling of the empirically observed cross-sections of estimated excess returns to approximate Φ by $\hat{\Phi}$, and thus obtains an approximation of $\Psi(\Phi)$ by $\Psi(\hat{\Phi})$. Bootstrap comparisons complement inferences on zero means that rely on the asymptotic normality of the t-tests since: (i) p-values based on the bootstrap distribution of coefficients are often more accurate in small samples when the unknown distributions are not Gaussian; and (ii) the cross-correlations within each cross-section are accounted for by construction.

[21] Brown *et al.* (1992) point out survivorship biases within a model of fund deletion where funds with higher residual variances get deleted disproportionately since they are more likely to encounter draws in the lower tail. Such a mechanism implies that the mean performance of terminated funds during the period when they survived should have a sample mean that is higher than the survivor funds on average, especially if we drop the last period of bad outcome for the fund that gets terminated. With our sample of real funds, however, even when we drop each terminating fund's performance during its last year of existence (when unfortunate outcomes in the tail may have been the termination-triggering event), we do not observe such behavior of sample means.

Table VII

Results From Survivorship-Biased Subsamples

The overall sample period is 1974Q1–1988Q4. Panel A shows two types of average performance percentile ranks for survivorship-biased subsamples. Let π_{it} be the percentile performance rank in quarter t of fund i belonging to subsample S relative to all the other funds in the entire sample. The two types of average percentile rank for funds considered are: $\pi_S^j = \sum_{i \in S} \sum_{t \in T_i^j} \pi_{it}$; $j = 1, 2$. For $j = 1$, T_i^1 includes all the quarters for which valid data on fund i is available; and for $j = 2$, T_i^2 includes the four quarters prior to and including the last quarter of data available for fund i. In addition to survivor's subsamples, results are shown from the subsample of funds that adopt a sales load.

Panel B reports on the persistence estimates of survivorship-biased subsamples following the method used in Table I, which relies on coefficient averages from stacked cross-sectional regressions. The proxy for the residual return is the residual for the market model regression with the EWNYSE benchmark.

Panel C reports on the performance of rank portfolios from survivorship-biased samples following the method used in Table III. For parsimony, only comparisons with a four-quarter evaluation period using the EWNYSE and the EWMF benchmarks are shown; other comparisons, not shown, are similar.

Panel A. Performance of Subsamples With Differing Survivorship Conditions.

Type of Subsample	Number of Funds	π_s^1 (Overall Performance Percentile)	π_S^2 (Performance Percentile in Last Four Quarters)
All	165	46	48
Survivors	94	51	not meaningful
Since 1974	51	51	not meaningful
Included post-1974	43	50	not meaningful
Nonsurvivors	71	41	45
Terminated	14	23	37
Merged	28	36	42
Switch to load	22	52	56
Other	7	60	47

Table VII—*Continued*

Panel B. Persistence Estimates in Survivorship-Biased Subsamples.

	Time-Average of Cross-Sectional Coefficients (t-statistics in parentheses; for sample of "Survivors and available since 1974", bootstrap-based "p-values" of the coefficient estimates are shown in brackets)								
	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{a}_4$	$\hat{a}_5$	$\hat{a}_6$	$\hat{a}_7$	$\hat{a}_8$	$\Sigma\hat{a}_i$
Entire sample (from Table I)	0.05 (1.03)	0.10 (2.56)	0.06 (1.62)	0.11 (2.88)	−0.05 (−1.69)	−0.05 (−1.84)	−0.06 (−1.79)	0.04 (1.61)	0.18
Survivors	0.05 (1.08)	0.11 (2.83)	0.06 (1.78)	0.11 (2.66)	−0.04 (−1.42)	−0.07 (−2.38)	−0.05 (−1.46)	0.03 (1.20)	0.19
Survivors and available since 1974	0.05 (1.03) [0.07]	0.10 (2.56) [0.00]	0.06 (1.62) [0.01]	0.11 (2.88) [0.00]	−0.05 (−1.69) [0.80]	−0.05 (−1.84) [0.98]	−0.06 (−1.79) [0.79]	0.04 (1.61) [0.02]	0.18 [0.00]
p-Value for F-test of coefficient equality across three samples	0.88	0.26	0.81	0.92	0.61	0.06	0.92	0.19	

Table VII—*Continued*

Panel C. Performance of Rank Portfolios in Survivorship-Biased Samples										
	Results on Rank Portfolios[a]								Best-Worst	Spearman's
	1	2	3	4	5	6	7	8	Portfolio[b]	Statistic[b,c]
Mean excess return										
Entire sample[d]	0.99	1.68	1.40	1.82	1.90	2.32	2.94	3.47		
Survivors' sample	1.13	1.61	1.72	1.75	2.70	2.65	3.34	3.23		
					Jensen's α					
EWNYSE benchmark										
Entire sample[d]	−1.85	−1.09	−1.33	−0.94	−0.89	−0.51	−0.10	0.14	1.99	2
	(−3.54)	(−3.06)	(−3.63)	(−2.57)	(−2.35)	(−1.26)	(−0.24)	(0.24)	[0.01]	[0.00]
Survivors and	−2.08	−1.08	−1.17	−0.98	−0.12	−0.18	0.11	−0.29	1.79	16
available since 1974	(−3.13)	(−2.44)	(−2.90)	(−2.19)	(−0.32)	(−0.35)	(0.22)	(−0.45)	[0.06]	[0.01]
EWMF benchmark										
Entire sample[d]	−0.92	−0.31	−0.56	−0.19	−0.13	0.25	0.70	0.98	1.90	2
	(−1.78)	(−1.48)	(−2.88)	(−1.15)	(−0.72)	(1.36)	(2.96)	(2.56)	[0.01]	[0.00]
Survivors and	−1.01	−0.37	−0.33	−0.26	0.65	0.58	0.96	0.62	1.63	10
available since 1974	(−1.57)	(−0.37)	(−0.33)	(−0.26)	(2.80)	(1.75)	(2.92)	(1.29)	[0.07]	[0.00]

[a]White's z-statistic, which corrects for the heteroscedasticity of the rank portfolio returns, is reported in parentheses.

[b]The p-values of observing the estimated statistic under the null hypothesis of zero predictability is reported in brackets.

[c]Spearman's statistic is computed as the sum of the squared differences between the octile's construction rank and its overall post-construction α-rank. It measures the predictability of α-ranks.

[d]For convenient comparison, we repeat results on "Entire sample" from Table III, Panel C.

For each bootstrap simulation, we construct a fake data set that mimics what would be available under the null hypothesis of no persistence while preserving the cross-sectional dependencies in excess returns. Each simulation begins by randomly drawing nine quarterly cross-sections of excess returns *without* replacement from the original sample of sixty quarters of data. These fill up locations one through nine in the simulated data set. For location ten in the simulated data set, we randomly draw a cross-section from the 52 quarters that do not occupy a position in the immediately previous eight locations (locations two through nine). We repeat this to fill up sixty locations. This particular method of constructing the data set ensured that each cross-sectional regression corresponding to equation (2) has full rank. Corresponding to row 2 of Panel B, Table VII, the means of the persistence coefficients across the 52 cross-sectional regressions for the fake data set are computed. We carry out 5000 bootstrap simulations.

Relative to the bootstrap distributions, the p-values for the original individual persistence coefficients are reported in brackets in Panel B of Table VII. The bootstrap-based p-value for each of the first four significant persistence coefficients is smaller than that obtained from comparison with the asymptotically Gaussian distribution of the t-test.[22] The bootstrap p-value for the sum of the first four persistence coefficients (sum = 0.32) is zero; the disappearance of persistence beyond four quarterly lags is confirmed by an insignificant bootstrap p-value of 81% for the sum of the persistence for lags 5 through 8 (sum = -0.12).

In Panel C of Table VII, we report some performance results of a representative hot hands strategy (four-quarter evaluation) with the subsample of 51 funds available for the entire 1974–88 period. Our earlier results with the entire sample are repeated from Table III for comparison. If anything, the persistence results are less clear-cut with the survivors' subsample: for example, (i) the Spearman statistic for α-rank ordering is less compelling (≥ 10), though statistically significant, when compared to its value with the entire sample (2), and (ii) the best-worst portfolio indicates a smaller maximum gain than with the entire sample.

In the case of octile performance categories, if the persistence is due solely to survivorship bias, which follows a mechanism similar to that conjectured by Brown *et al.* (1992), the pattern relating octile rank to performance rank will manifest a U-shape rather than a simple spurious monotonic relation. Within the Brown *et al.* framework, a sufficient condition for a U-shaped performance pattern is that two or more ranks get located below the uncondi-

[22] In our setup, we expect to find that the asymptotic t-test will understate the significance of the coefficients since the variables used in estimating equation (2) are estimated excess returns that sum to zero over time by construction. For a typical cross-sectional regression, when the left-hand side variable is above (below) zero, the expected value (by construction) for the right-hand side variables will be below (above) zero implying a downward bias of $1/(N-1)$ in the coefficient estimates. Under the null hypothesis of zero coefficients, the expected estimate will be $-1/(N-1) \approx -0.02$ with our sample size of sixty observations per fund, which value is indeed observed in the bootstrap simulations.

tional mean of funds' performance return.[23] This is very likely with relatively fine gradations of performance categories. For instance, this condition is met for the simulation parameters used by Brown *et al.* if we sort funds into octiles instead of their two categories. Repeating their simulations with octile categories, we have verified the U-shaped pattern. In our empirical results with octile categories, we do not observe a U-shaped pattern; instead, we observe a clear monotonic relation that suggests true performance persistence rather than a survivorship-related effect.

Overall, the results in Table VII suggest that survivorship bias is probably not an important issue for performance studies with typical mutual fund samples.

IV. Conclusion

During the 1975–88 period, substantial gains were available from investing in the mutual fund equivalents of last year's pennant winners. Specifically, no-load growth-oriented mutual funds that performed well relative to their brethren in the most recent year continue to be superior performers in the near term (one to eight quarters). A strategy of selecting, every quarter, the top peformers based on the last four quarters (such as the top octile) can significantly outperform the average mutual fund, albeit doing only marginally better than some benchmark market indices. Icy hands, the evil counterpart of hot hands, also shows up in our sample: funds that perform poorly in the most recent year continue to be inferior performers in the near term. Indeed, they are more inferior than hot hands are superior. While there is little support for funds that are sustained superior performers, we do identify some funds that are sustained underperformers.

The hot hands phenomenon does not appear to be driven by already known anomalies, since superior performance is also achieved relative to an eight-portfolio benchmark that accounts for effects of firm size, dividend yields, and reversion in returns. Our sample was carefully constructed to avoid problems of survivorship bias. As a practical matter, we find from subsample analysis that any such bias appears to be unimportant for studying persistence in mutual fund performance. The benefits from hot hands strategies are verified in an independent sample, which has been previously studied in the literature, and also appear in 1989–90. Future research should seek to explain the factors that underlie the discovered patterns of time decay in the performance of funds.

[23]See Hendricks, Patel, and Zeckhauser (1992). Briefly, the U-shape arises because funds in the lowest octile have disproportionately higher variances. Contingent on surviving, they will have higher means than the second-octile funds, hence the left leg of the U. Here we assume that the actual performance of the lowest two octiles in the construction quarter was below the unconditional mean. Further, since the funds with the worst outcomes are not included in a sample with survivorship bias, this left leg of the U-shaped pattern will be lower than the right leg.

G. P. Szego and K. Shell, eds.: *Mathematical Models in Investment and Finance* (North Holland, Amsterdam).

Treynor, J., and F. Mazuy, 1966, Can mutual funds outguess the market?, *Harvard Business Review* 44, 131–136.

White, H., 1980, A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica* 48, 817–838.

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