ELSEVIER Journal of Financial Economics 54 (1999) 337–374

JOURNAL OF Financial ECONOMICS

www.elsevier.com/locate/econbase

Survivorship bias and attrition effects in measures of performance persistence☆

Jennifer N. Carpenter, Anthony W. Lynch*

Department of Finance, Stern School of Business, New York University, New York, NY 10012-1126, USA

Received 3 August 1998; received in revised form 22 October 1998; accepted 12 August 1999

Abstract

We simulate standard tests of performance persistence using alternative return-generating processes, survival criteria, and test methodologies. When survival depends on performance over several periods, survivorship bias induces spurious reversals, despite the presence of cross-sectional heteroskedasticity in performance. Look-ahead biased methodologies and missing final returns typical of U.S. mutual fund datasets can also materially affect persistence measures. Our results reinforce previous findings that U.S. mutual fund performance is truly persistent. When fund performance is truly persistent, fund attrition affects persistence measures, even when the sample includes all nonsurvivor returns. We also examine the specification and power of the various persistence tests. © 1999 Elsevier Science S.A. All rights reserved.

PACS: JEL classification: G23; C15

Keywords: Survivorship bias; Attrition; Persistence; Reversals; Mutual funds

☆We would like to thank Stephen Brown, Edwin Elton, Wayne Ferson, Martin Gruber, Bruce Grundy, William Schwert (the editor), and an anonymous referee for helpful comments and suggestions. We also acknowledge the research assistance of Olesya Grishchenko.

* Corresponding author: Tel.: + 1-212-998-0352; fax: + 1-212-998-0350.

E-mail address: alynch@stern.nyu.edu (A.W. Lynch)

PII: S0304-405X(99)00040-9

1. Introduction

This paper examines the impact of survivorship bias and attrition on measures of performance persistence, which typically examine performance during a ranking period and a subsequent evaluation period. We simulate standard persistence tests using samples calibrated to match the history of the U.S. mutual fund industry under various assumptions about return-generating processes, survival criteria, and data availability. We have five main findings. First, despite the presence of cross-sectional heteroskedasticity in fund alphas, when survival depends on performance over several periods, survivorship bias creates a reversal effect which dominates the persistence effect discovered by Brown et al. (1992). Second, even when samples include all nonsurvivors, look-ahead biased methodologies (which require funds to survive a minimum period of time after a ranking period) and missing final-period returns still materially bias statistics. Third, when death rates are calibrated to mutual fund data, survivorship bias cannot explain the degree or pattern of performance persistence documented empirically.

Fourth, when the return-generating process has true persistence, we find that the attrition, or disappearance, of poor performers creates an effect that is distinct from survivorship bias. In particular, even when all fund returns are included in the persistence tests, the systematic disappearance of poor-performing funds causes the values of persistence measures to differ from the values that would be found in samples with no attrition. Finally, the comparison of simulated rejection frequencies for various persistence tests suggests that the t-test for the difference between top and bottom portfolios ranked by past performance is the best specified under the null hypothesis of no persistence and one of the most powerful against the alternatives we consider.

Performance persistence is an important issue in a number of contexts. The question of whether mutual fund performance persists is crucial in explaining how investors should choose funds. The track records of fund managers are only important to consider in choosing among competing funds if performance is persistent. Similarly, the issue of persistence or 'momentum' in stock returns has important implications for trading strategies of investors.

Performance persistence also has important implications about the nature of markets. According to Gruber (1996), the persistence of good mutual fund performance suggests that some managers have superior ability and explains why actively managed funds prevail despite evidence that the average fund underperforms passive benchmarks. On the other hand, persistence in stock returns might imply that information in past returns can be used to generate abnormal profits, which would contradict the efficient markets hypothesis.

Unfortunately, the fact that poor performers tend to disappear can obscure the empirical estimation of the degree of persistence. One problem is that samples that do not include all returns of disappearing funds introduce survivorship bias in performance measures. Most recent mutual fund studies strive to eliminate such bias by incorporating all available data on fund returns. Nevertheless, some degree of survivorship bias is unavoidable if complete data on disappearing funds are difficult to find. This problem is of special concern with new hedge fund samples. For example, Ackermann et al. (1996) and Brown et al. (1997) note that their hedge fund samples may not include some defunct funds. Further, Brown et al. (1997) use annual returns and note that disappearing funds can be missing up to a year of final returns. A second problem is that, even when samples include all available data, some methodologies introduce survivorship bias by imposing minimum survival requirements.

A final problem is that, even without survivorship bias, the mere attrition of poor performers alters persistence measures because it changes the composition of the sample. The direction of the attrition effect depends on the nature of the persistence. When using a sample with attrition, a researcher must take this effect into account in order to obtain unbiased parameter estimates for the true return-generating process.

One line of research suggests that survivorship imparts an upward bias to persistence measures. In particular, Brown et al. (1992), hereafter BGIR, demonstrate that if fund volatility is constant across time but varies cross-sectionally, and if funds disappear each period according to whether or not their performance that period falls into the bottom fraction of funds, then survivorship induces spurious persistence. Conditional on making the cut, the best funds tend to have high volatility. Conversely, high-volatility funds have the highest expected performance, conditional on survival. Therefore, in a sample of survivors, first-period winners tend to have high volatility and subsequently win in the second period. On the other hand, a number of papers, including BGIR, Grinblatt and Titman (1992), and Hendricks et al. (1993), argue that if fund survival depends on average performance over several periods, then survivorship induces spurious reversals. First-period losers must subsequently win in order to survive.

A number of persistence studies examine survivorship bias empirically by comparing test results for a survivor-only sample with those for a full sample. Both Hendricks et al. (1993) and Carhart (1997) find that persistence is weaker in the sample of survivors. Carhart (1997) distinguishes two effects. Survivorship bias is a property of the sample selection method. It results when the sample includes only funds that survive until the end of the sample period. By contrast, look-ahead bias is a property of the test methodology. Carhart (1997) uses persistence measures based on the returns to investment strategies that annually sort funds into ten portfolios ranked on past performance. His full sample keeps funds in their portfolios until they disappear, and then rebalances the portfolio. His look-ahead biased sample requires funds to survive a minimum period of time after the ranking period in order to enter into the portfolio strategy. His survivor-biased sample includes only funds that survive until the end of the sample period. He finds the strongest persistence in the full sample and the weakest in the survivor-biased sample.

While empirical evidence on survivorship bias in specific samples is important, our paper makes an incremental contribution. By examining survivorship bias in a controlled setting, we provide a general understanding of the exact effect of different combinations of return processes, survival criteria, data availability, sample selection, and test methodologies. We are also able to characterize the size and power of the persistence tests. In addition, our paper shows how fund attrition alters persistence measures, which is not possible to do empirically because all mutual fund samples suffer from attrition. Finally, our paper has implications about performance persistence tests in any setting in which poor performers systematically disappear. Although we focus on the case of mutual funds, our results potentially apply to studies of persistence in stock returns.

In calibrating our simulations, we focus mainly on the characteristics of the sample of Carhart (1997). There are two reasons for this. First, it is one of the most comprehensive mutual fund data sets in the literature, including new funds as well as seasoned.[1] Second, Carhart's data set is soon to become available from the Center for Research in Security Prices (CRSP). We generate 33-year samples of fund alphas with three different return-generating processes. One process has no true persistence. In that case, fund alphas are independent with mean zero, but are cross-sectionally heteroskedastic. The cross-sectional distribution of fund volatility is the same as that in the BGIR simulations, chosen to match characteristics of the sample of Goetzmann and Ibbotson (1994). The alternative generating processes introduce true persistence either with independent alphas with cross-sectionally heterogeneous means, reflecting permanent differences in ability, or else with zero-mean alphas that follow moving average processes, a 'hot-hands' phenomenon.

In each case, we consider two different survival criteria. The single-year criterion cuts funds based on their performance over the preceding year, while the multiyear criterion considers average fund performance over the preceding five years. These alternatives seem to bracket the true survival criteria in the mutual fund data, based on evidence about final-period returns of disappearing funds in Carhart (1997) and Brown and Goetzmann (1995). Our results are also likely to be indicative of biases in hedge fund studies, since Brown et al. (1999) suggest that hedge fund survival can depend on performance over periods as short as one or two years.

The growth rate of the number of funds and the fund attrition rate in the simulated samples match those of the Carhart (1997) sample. For the sake of

[1] Other samples that include defunct funds are available, such as Malkiel (1995), Brown and Goetzmann (1995), and Elton et al. (1996b). While some of these samples have fewer missing returns than Carhart's sample, they are much less comprehensive.

comparison, we also generate samples with no attrition. Finally, we consider the effect of missing returns. Although the Carhart (1997) data set is uniquely comprehensive and complete, even that data set is missing an average of two months of final returns for disappearing funds.[2] With each sample, we consider two possibilities: either all data are available or the final two months of returns of disappearing funds are missing.

With each sample, we test for persistence using methodologies commonly found in the literature. These include tests measuring the performance of decile portfolios ranked on past performance, contingency table tests, and cross-sectional regressions of current alphas on past alphas. One main result is that when survival depends on multi-period performance, the reversal effect dominates the BGIR persistence effect, even though fund returns are heteroskedastic. Missing data alone bias mean differences in alphas of top and bottom decile portfolios downward as much as 0.26% per year in the sample with no true persistence. Look-ahead bias and survivorship bias further reduce the mean performance difference by as much as 1.27% per year. These biases are even larger when persistence is present. This downward bias is consistent with empirical findings, as in Hendricks et al. (1993) and Carhart (1997), that persistence is stronger in full samples than in survivor-only samples.

In the simulations without true persistence in fund performance, even the single-period survival criterion cannot create enough spurious persistence to match the magnitudes found by Carhart (1997) and others. For example, in the simulated full sample with no true persistence and with the single-period survival criterion, the missing data bias creates a difference between the average alphas of top and bottom decile portfolios of 0.14% per year, using three-year ranking periods and one-year evaluation periods. By contrast, in the full sample of Carhart (1997), the corresponding performance difference is 0.36% per month. Similarly, the simulated look-ahead biased performance difference is 0.58% per year while Carhart finds 0.33% per month, and the simulated performance difference for the survivor-biased sample is 0.16% per year compared with 0.18% per month in the Carhart sample. Taken together, our results confirm that the mutual fund performance persistence documented by empirical studies is not spurious, but rather is a feature of the true data-generating process.

Our results also have qualitative implications about the degree of persistence in stock returns. Studies of stock return reversals and momentum typically use the CRSP database. Shumway (1997) reports that a significant number of delisting returns are missing from CRSP. Moreover, the missing returns are systematically associated with poor firm performance. The poor performance events that trigger delisting, such as bankruptcy or failure to meet capital requirements, are of a multiperiod nature. Our simulations demonstrate that

---

[2] We thank Mark Carhart for this information.

when the survival criterion is multiperiod, missing final-period returns bias persistence measures towards reversals.

The results of Shumway (1997) confirm this prediction. He finds that augmenting the CRSP data set with delisting returns from OTC data substantially reduces the reversal effect documented by DeBondt and Thaler (1985). Like mutual fund returns, stock returns display cross-sectional heteroskedasticity, so we would expect the BGIR persistence effect to be at work here as well. Yet Shumway's results show that the multiperiod reversal effect dominates the BGIR persistence effect in the case of stocks, just as it does in our simulations.

Given Shumway's results, our paper has implications for studies that document persistence in stock returns, such as Jegadeesh and Titman (1993) and Fama and French (1996). In particular, our simulations suggest that the spurious reversals induced by multiperiod survival criteria are robust to the presence of true persistence in the data. Thus, the missing delisting returns on CRSP have probably caused these studies to understate the magnitude of the persistence.

Estimating the nature and degree of the persistence in returns is important both for researchers trying to understand the nature of markets and for practitioners developing investment strategies. The simulations with true performance persistence show that even with samples and methodologies that are completely free of survivorship bias, the mere attrition of funds alters persistence measures, because it changes the cross-sectional composition of funds. For example, under the specification with heterogeneous means in performance, attrition removes low-mean funds, which narrows the dispersion in ability and reduces persistence measures. Our results suggest that estimation of theoretical models of persistence in fund or stock returns must properly account for the impact of attrition in the data in order to lead to correct inferences. Further, the impact of attrition on persistence measures depends on the nature of the persistence.

Finally, to illustrate the specification and power of the different persistence tests, the simulations yield upper and lower test rejection frequencies. In general, both the *t*-test for the difference between the top- and bottom-ranked portfolios without overlapping evaluation periods and the chi-squared test on counts of winners and losers are well specified and powerful against the alternatives we consider. The difference *t*-test is more powerful, while the chi-squared test is more robust to the presence of survivorship bias. The Spearman test is very powerful but overrejects under the null. The *t*-test for the slope coefficient in the cross-sectional regression of current performance on past performance is neither well specified nor powerful. The test for spurious persistence proposed by Hendricks et al. (1997) performs well in samples with no true persistence, but can indicate spurious persistence in samples with both true persistence and survivor bias.

The paper proceeds as follows. Section 2 reviews the literature on persistence in mutual fund performance. Section 3 describes the simulations. Section 4 presents the results, and Section 5 concludes.

2. Related literature

While a number of articles examine the effects of survivor bias on measures of average mutual fund performance, only a few focus on the effects of survivor bias on measures of persistence. BGIR show that when fund performance is cross-sectionally heteroskedastic and survival depends on single-period performance, conditioning on survival causes two-period contingency table tests to indicate performance persistence when none exists. In the sample of survivors, funds in the upper half of performers in the first period tend to remain in the upper half in the second period. Hendricks et al. (1997) extend this work to show that, conditional on survival, expected second-period performance is a J-shaped function of first-period performance.

Other papers argue that if survival depends on multiperiod performance, then conditioning on survival should bias persistence measures towards reversals. This idea appears in the appendix of BGIR, as well as in Grinblatt and Titman (1992, fn. 4) and Hendricks et al. (1993). Empirical comparisons of persistence tests on full and survivor-biased samples, such as Hendricks et al. (1993) and Carhart (1997), show that survivor bias weakens persistence measures. This suggests that the survival of real mutual funds depends at least partly on performance over several periods.

Brown and Goetzmann (1995) and Carhart (1997) provide direct evidence that survival criteria for mutual funds are multiperiod. Brown and Goetzmann (1995) find that the probability of fund disappearance depends on past returns as far as three years back. Carhart (1997) finds that nonsurvivors underperform for as many as five years prior to disappearance. He shows that survivor bias in average fund returns grows with the length of the sample period and proves that this phenomenon occurs with a multiperiod survival criterion, but not with a single-period criterion. In a sample that does not add new funds, Elton et al. (1996b) also find that survivor bias in average fund returns grows with the length of the sample period. The evidence of Brown et al. (1999) suggests that survival criteria for hedge funds are shorter than five years.

In general, empirical studies find that mutual fund performance is persistent. In addition to the articles mentioned above, these studies include Goetzmann and Ibbotson (1994), Malkiel (1995), Elton et al. (1996a), Gruber (1996), Carhart (1998), and Wermers (1997). By contrast, Brown et al. (1999) find no persistence in hedge fund performance.

A related literature examines persistence in stock returns. Jegadeesh and Titman (1993) and Fama and French (1996), among others, find a so-called 'momentum effect' in short-term stock returns, while DeBondt and Thaler (1985) document reversals over longer horizons.

Other papers that use simulation to study mutual fund performance include Kothari and Warner (1997) and Wermers (1997). Kothari and Warner (1997) examine the properties of common mutual fund performance measures. Rather than generate fund returns from a parametrized distribution, their simulations form funds from random samples of NYSE and AMEX securities. They show that standard performance measures indicate abnormal performance and market-timing ability when none exists. Thus, the focus of their paper is on average performance, even in the absence of attrition, while our focus is on performance persistence in the presence of attrition. Wermers (1997) simulates fund performance in a setting where stock returns follow an AR(1) process and funds follow either contrarian or momentum investment strategies. Although his empirical work examines performance persistence, his simulations focus only on survivor bias in average performance.

3. Simulations of performance studies

Each of our simulations generates 33 years of performance measures, or alphas, for a collection of fund managers, and then conducts tests for performance persistence using the simulated sample. To match the U.S. mutual fund industry, we set the initial number of funds equal to 213, the number of funds in the sample of Carhart (1997) in 1962, and the annual growth rate of the number of funds equal to 5.8%, the geometric mean growth rate of the number of funds in Carhart's sample. The simulations use a variety of return-generating processes, survival criteria, sample selection methods, and persistence tests in different combinations. The remainder of this section describes these choices in detail.

3.1. Alpha processes

We consider three alternative generating processes for the alphas of fund managers. In every case, alphas are cross-sectionally independent and heteroskedastic. The variance of manager i's annual alpha, σ_i^2, has the cross-sectional distribution used by BGIR:

$$\sigma_i^2 = 0.05349(1 - \beta_i)^2, \tag{1}$$

where

$$\beta_i \sim \mathcal{N}(0.95, 0.25^2). \tag{2}$$

3.1.1. Independent zero-mean alphas

Under the first specification, IND, used by BGIR, alphas of each manager are independent across time with mean zero. Letting $\alpha_{i,t}$ denote the alpha of manager i in year t,

$$\alpha_{i,t} \sim \mathcal{N}(0, \sigma_i^2). \tag{3}$$

Under the IND specification, the alpha processes have no true persistence. By contrast, under the other two specifications, performance is persistent.

3.1.2. Independent alphas with cross-sectionally heterogeneous means

Under the HET specification, fund returns are intertemporally independent, but managers have permanent differences in mean returns. Such a situation might arise if fund expenses are constant across time but differ cross-sectionally. Formally,

$$\alpha_{i,t} \sim \mathcal{N}(\mu_i, \sigma_i^2), \tag{4}$$

where

$$\mu_i \sim \mathcal{N}(0, \sigma_{\mu,i}^2), \tag{5}$$

$$\sigma_{\mu,i} = \min(0.02, 0.99\sigma_i). \tag{6}$$

Most of the time $\sigma_{\mu,i}$ is 0.02, which is a conservative estimate of the dispersion in managerial ability in the data. In particular, Carhart (1997) finds a difference in average returns of 0.63% per month between the bottom and top deciles of fund managers, ranked by past one-year returns. We use this figure to calibrate the cross-sectional volatility of mean performance in the simulations. As the results in Section 4 show, the simulated difference in performance for the top and bottom deciles is less than that found by Carhart (1997) for both returns and alphas. Thus, our calibration yields a lower degree of persistence than that in the data.[3]

The existence of a comparable MA(1) process described below requires that the volatility of μ_i remain below σ_i. For this reason, we set $\sigma_{\mu,i} = \min(0.02, 0.99\sigma_i)$. As a result, managers with betas near one will have both low volatility and mean alphas near zero, which seems reasonable.

3.1.3. MA(1) alphas with zero-mean

An alternative way of modeling persistence in performance incorporates autocorrelation in fund performance, the 'hot hands' property described by Hendricks et al. (1993). Under the MA1 specification, each manager's performance follows a zero-mean moving average process of order one, calibrated to

[3] If the deciles in Carhart (1997) were sorted by true mean, a spread in average evaluation period performance of 0.63%/month would imply an annual cross-sectional volatility in true means of $12 \times 0.63\%/(2 \times 1.64) \approx 2\%$, using the 0.05 and 0.95 fractiles of the normal distribution to approximate the mean performance of the bottom and top deciles. Because sorting on past one-year performance does not perfectly sort on true mean, this method of estimating the dispersion in ability understates the actual level in the data. On the other hand, using returns rather than alphas to calibrate persistence increases the estimate of dispersion in abnormal performance because part of the true dispersion in mean returns is due to dispersion in betas.

make the variance of $\alpha_{i,t}$ and the cross-sectional covariance of $\alpha_{i,t}$ with $\alpha_{i,t-1}$ the same as under the HET specification. In particular, given σ_i for manager i, we choose the MA coefficient θ_i and the white noise volatility $\sigma_{w,i}$ so that var($\alpha_{i,t}$) and E($\alpha_{i,t}\alpha_{i,t-1}$) are the same as under the HET specification:

$$\alpha_{i,t} = w_{i,t} + \theta_i w_{i,t-1}, \tag{7}$$

where

$$w_{i,t} \sim \mathcal{N}(0, \sigma^2_{w,i}), \tag{8}$$

$$\sigma^2_{w,i} = \sigma^2_{\mu,i}(\gamma_i + \sqrt{\gamma_i^2 - 4})/2, \tag{9}$$

$$\theta_i = (\gamma_i - \sqrt{\gamma_i^2 - 4})/2, \tag{10}$$

$$\gamma_i = 1 + \sigma_i^2/\sigma^2_{\mu,i}. \tag{11}$$

This construction creates a negative relation between σ_i and θ_i. Alternatively, we could consider only those managers with the average σ and fix θ at a value that equates the cross-sectional covariance of α_t with α_{t-1} under the HET and MA1 specifications, but then the cross-sectional covariance of α_t with α_{t-1} across all managers would differ under the two specifications. Since the goal is to assess how true persistence affects persistence measures, our choice seems reasonable.

3.2. Survival criteria

Carhart (1997) finds that 3.6% of funds disappear from his sample in an average year. Accordingly, our simulations eliminate the 3.6% poorest-performing funds each year, based on one of two alternative measures of past performance. The simulations using the single-year criterion, SY, cut funds based on their past year's alpha. The simulations using the multiyear criterion, MY, cut funds based on their average alpha over the last five years, as suggested by Fig. 1 of Carhart (1997). At any time there are up to five cohorts: one-year old funds, two-year olds, three-year olds, four-year olds, and those at least five years old. The SY simulations cut the bottom 3.6% of each cohort. The MY simulations cut the poorest of funds at least five years old; the number cut is equal to 3.6% of the sample. Since attrition under the MY simulations does not begin until year 5, the performance studies consider a 29-year sample period running from year 5 to year 33. We also tried a hybrid criterion that cuts 0.8% of each cohort based on past one-year alphas and then cuts from the cohort of funds at least five years old based on their five-year alphas, such that the total number cut represents 3.6% of the sample. The results for this hybrid criterion (unreported) lie between those for the SY and MY cases and are typically closer to those for the MY case.

To examine the effect of attrition when persistence does exist, we also generate control samples with no attrition, labeled Z. Finally, we generate a smaller sample with no attrition, labeled ZS, with an initial number of funds equal to 79.

Comparing test results from the Z and ZS samples is useful for judging the extent to which differences in test statistics between full and survivor-only samples are attributable merely to a change in sample size. The sizes of the Z and ZS samples represent upper and lower bounds on the sizes of all other samples in the performance studies.

3.3. Data availability and sample selection

We assume either that all data are available, ALL, or that the final two months of performance of disappearing managers are missing from the data set, MSS. In addition, the simulations construct two samples for the performance studies. The survivor-biased sample, SB, includes only funds with data available until the end of the sample period. The 'comprehensive' KU sample keeps funds as long as they have data available.

3.4. Persistence tests

With each simulated sample, we conduct standard tests for persistence in fund performance. One class of tests sorts funds into deciles each year based on past performance and measures the average difference in performance between the top and bottom deciles and the Spearman rank correlation. The other class of tests forms contingency tables that count the number of funds that were winners, relative to median performance, in both a ranking period and an evaluation period, and then measures the cross-product ratio and its significance. Related tests regress evaluation-period performance on ranking-period performance in the cross-section. We use 2500 replications to simulate the mean values of the persistence measures and test statistics, as well as upper and lower rejection rates.

3.4.1. Performance-ranked portfolio strategies

Most recent studies of performance persistence in mutual funds examine returns to portfolios of funds sorted by past performance.[4] The simulations sort funds each year into decile portfolios based on their average alpha over the preceding ranking period and then measure the equally weighted average portfolio alpha over the subsequent evaluation period. We consider three combinations, 1R1E, 3R1E, and 3R3E, with, for example, 3R1E indicating a three-year ranking period and a one-year evaluation period. Elton et al. (1996a), Gruber (1996), and Carhart (1997) all use one or more of these combinations.

[4] Examples include Hendricks et al. (1993), Brown and Goetzmann (1995), Elton et al. (1996a), Gruber (1996), Carhart (1997, 1998), and Wermers (1997). Similarly, studies of persistence in stock returns using this methodology include DeBondt and Thaler (1985), Jegadeesh and Titman (1993), and Fama and French (1996), among others.

The decile portfolios include only funds with alphas available throughout the entire ranking period. With the survivor-biased (SB) samples, all funds in a given portfolio survive throughout the entire evaluation period, so there is no question of rebalancing the portfolio when funds disappear. With the comprehensive (KU) samples, some funds disappear, or their data become unavailable, before the end of the sample. We treat this occurrence in one of three alternative ways.

The include-available-alphas (IAA) methodology keeps funds in a portfolio as long as data are available, and then rebalances the portfolio equally among the remaining funds. This captures the idea of the 'follow the money' methodology of Elton et al. (1996a,b) and Gruber (1996). In the missing data (MSS) case, in which the last two months of a disappearing fund's data are unavailable, we assume that a fund's annualized alpha in the last two months of the year is the same as in the first ten. If we generate returns monthly, instead of annually, then the first ten and last two months of the year would have different average alphas. However, in the i.i.d. cases, their means would be the same, both unconditionally and conditional on survivorship, as long as we cut funds based on annual performance, so the simplification using annual alphas imparts no bias. In the MA case, the impact using annual alphas instead of monthly alphas depends on the monthly MA specification. To the extent that average monthly returns for real funds are worse in the final two months than in the final year as a whole, our model understates the effect of the missing returns.

Two other methodologies introduce what Carhart (1997) calls 'look-ahead' bias by eliminating funds that fail to survive a minimum period of time after the ranking period. The length of this minimum period is equal to the length of the ranking period. The partial look-ahead (PLA) methodology first forms the deciles using all funds with valid ranking-period returns, and then eliminates the funds that disappear before the end of the evaluation period. By contrast, the full look-ahead (FLA) methodology first eliminates the disappearing funds, and then forms the decile portfolios. Thus, we decompose the total look-ahead bias into two effects, one from imposing the minimum survival requirement and one from changing the decile breakpoints. The FLA methodology is the one Carhart (1997) labels 'look-ahead biased'.

With regard to contingency tables that define 'winner' and 'loser' funds relative to the sample median firm and count the number of winner-winners during a ranking and subsequent evaluation period, most studies essentially use the FLA methodology. They first eliminate funds that fail to survive both periods, and then define winners and losers in each period relative to the median survivor's performance. By contrast, the contingency tables of Brown and Goetzmann (1995), which include the categories winner-gone and loser-gone, essentially use a PLA methodology, because they define first-period winners and losers relative to the full sample of funds extant in the first period.

Each simulation generates an annual time series of annualized evaluation-period alphas for each of the ten decile portfolios. In the 3R3E case, the three-year evaluation-period alphas overlap. In the 1R1E case, the time series is 28 years long, with ranking periods from year 5 to year 32. In the 3R1E and 3R3E cases, the time series is 24 years long, with ranking periods from 5–7 to 28–30.

The time series of evaluation-period alphas yields a number of statistics. AV10 and AV1 are the average alphas of the portfolios of top and bottom past performers, respectively. DIF is AV10 minus AV1. TDIF is a t-statistic for DIF. When the evaluation period is one year, the standard error for DIF assumes that the evaluation-period alphas are independent. The 3R3E case uses a Newey–West standard error with two lags, which has an asymptotic justification. SPEAR is the Spearman rank correlation between ranking-period ranks (RPR) and evaluation-period ranks (EPR), where the evaluation-period rank of a given decile portfolio is based on its average evaluation-period alpha across time. N is the average number of test funds per year.

For the IND samples, we compute the mean values of these statistics across replications as well as upper and lower rejection frequencies for two-sided 5% tests. For the HET and MA1 samples, we compute rejection rates for one-sided 2.5% tests for persistence.

Finally, each replication computes the J-shape statistic developed by Hendricks et al. (1997), the t-statistic for the linear term in the regression of EPR on a constant, RPR, and RPR^2. A significantly positive statistic should indicate true persistence (a monotone positive relation), while a significantly negative t-statistic should indicate a spurious persistence caused by survivor bias (a J-shape). We report the rejection rates for these tests for true and spurious persistence as well.

3.4.2. Contingency tables

Several studies of mutual fund performance measure persistence using contingency tables that classify funds as winners or losers in each of two consecutive time periods and count the number of winner-winners (WW), winner-losers (WL), and so on. These include Malkiel (1995), Brown and Goetzmann (1995), and BGIR.

The simulations examine four different ranking period\evaluation period pairs during the history of fund returns. The pairs are 5\6, 32\33, 5–7\8–10, and 28–30\31–33, where, for example, 5–7\8–10 indicates that the ranking period is years 5 through 7 of the simulated history and the evaluation period is years 8 through 10. Whether the underlying data set is survivor-biased (SB) or keeps funds until they disappear (KU), a given contingency table considers only funds with data available through both the ranking and evaluation periods. A fund is a winner or a loser in a given period depending on whether its alpha is above or below the median performance of funds in this group. This classification is analogous to the FLA ranking for performance decile portfolios described earlier.

The number of funds in the contingency table, N, and the number of winner-winners, WW, determines the remaining three counts, WL, LW, and LL. They therefore determine the value of three common persistence measures: the cross-product ratio, $\mathrm{CP} = (\mathrm{WW} \times \mathrm{LL})/(\mathrm{WL} \times \mathrm{LW})$, a chi-squared statistic with one degree of freedom, $\mathrm{CHI} = ((\mathrm{WW} - N/4)^2 + (\mathrm{WL} - N/4)^2 + (\mathrm{LW} - N/4)^2 + (\mathrm{LL} - N/4)^2)/N$, and the percentage repeat winners, $\mathrm{PRW} = \mathrm{WW}/(N/2)$. The test for persistence uses the chi-squared statistic to measure significance, rejecting independence if CHI exceeds a critical value (3.84 for a 5% test). This defines upper and lower rejection regions for CP and PRW. We simulate mean values of the statistics as well as upper and lower rejection rates in a 5% test. We also simulate mean values of the average ranking-period alpha, ABAR1, the average evaluation-period alpha, ABAR2, and the average value of σ_i, SIGBAR, for funds in the contingency table sample.

Finally, like BGIR, we use the contingency table sample to simulate TCS, the t-statistic for the slope coefficient in the cross-sectional ordinary least squares (OLS) regression of evaluation-period alphas on ranking-period alphas. Grinblatt and Titman (1992) and Brown et al. (1998) use this statistic as a measure of performance persistence.

3.5. Decomposition of effects

To understand how the various simulations differ, it is helpful, where possible, to organize the samples to be increasingly exclusive and then to identify the incremental changes. First consider the performance deciles. For any alpha process, IND, HET, or MA1, the Z sample contains complete histories for all managers. Going from the Z sample to a sample with attrition, but completely free of survivor bias, ALL-KU-IAA, creates an 'attrition effect'. Persistence measures can change because cutting low performers alters the cross-sectional composition of managers in the sample. Going from ALL-KU-IAA to MSS-KU-IAA introduces a 'missing data bias' caused by replacing each disappearing fund's final two months of alphas with the average alpha of survivors in the same decile.

Under either the ALL or MSS assumptions about data availability, going from IAA to PLA produces the 'partial look-ahead bias' caused by eliminating any partial evaluation-period alphas of funds that disappear during the evaluation period. Going from PLA to FLA creates a 'ranking effect' because the decile break points change when we eliminate the disappearing funds before the ranking. The total 'look-ahead bias' combines these last two effects. Going from FLA to the SB sample creates an 'incremental survivor bias' by eliminating funds that disappear after the evaluation period, typically high-volatility, low-mean funds.

Roughly analogous effects are present in the contingency table methodologies. For the cases with one-year ranking and evaluation periods, going from

Z to ALL-KU creates the attrition effect and going from ALL-KU to MSS-KU creates the missing data effect. For either the ALL or MSS cases, going from one-year to three-year ranking and evaluation periods introduces a look-ahead bias, while going from KU to SB introduces the incremental survivor bias.

4. Results

We begin this section with a discussion of the basic survivor biases, using the results for the IND specification as the main illustration. We then examine the survivor biases under the alternative specifications and compare the simulation results here with the empirical results of Carhart (1997). Next, we describe the effects of attrition in the cases with true persistence, and we conclude with a discussion of test specification and power.

4.1. Survivor biases

A number of the methodologies and samples introduce survivor bias in the persistence measures by requiring funds to survive beyond the evaluation period in order for all or part of their evaluation-period alphas to count. For example, in the MSS cases, only funds that survive until after the evaluation period count in the last two months of the evaluation period. In the look-ahead biased cases with three-year ranking periods, the first two years of alphas are survivor-biased, and in the SB sample, all alphas except those in year 33 in the ALL case are survivor-biased. This section describes the main effects of survivor bias on persistence measures.

4.1.1. Persistence effects of heteroskedasticity

The most basic effect is that identified by BGIR whereby cross-sectional heteroskedasticity in the alphas introduces spurious persistence in survivor-biased samples. The higher a fund's volatility, the greater is its expected alpha in any year, conditional on surviving that year's cut. Therefore, cases that require evaluation-period alphas to make the cut make high-volatility funds look better than low-volatility funds. To the extent that past winners tend to have higher volatility than past losers, this results in spurious performance persistence.

Past winners tend to have higher volatility than past losers for the following reason. High-volatility funds have extreme performance. Extremely good performers survive, but extremely poor performers disappear. Surviving losers have only moderate volatility. Therefore, in the subset of funds that survive the ranking period and enter into the evaluation period, the winners have higher volatility than the losers. More precisely, Hendricks et al. (1993, 1997) note that among survivors, average volatility is a J-shaped function of past performance; the closer is fund performance to the unconditional mean, the lower is the fund's expected volatility. Thus, the average decile volatility is a J-shaped function of decile rank, so the bottom decile has lower average volatility than the top, but the relation is not monotonic. With contingency tables, which divide the sample in half, winners have higher average volatility than the losers.

The cleanest examples of the BGIR persistence effect are in the case of independent zero-mean alphas with the single-year survival criterion (IND-SY), although the effect is present in the other cases to some degree. Tables 1 and 2 contain results for the IND specification. Table 1 presents the persistence measures for the performance-ranked portfolios and Table 2 presents the results for the contingency tables and cross-sectional regressions. Note first that the cases that do not condition evaluation-period alphas on survival (Z, ALL-KU-IAA, ALL-KU-5\6, and ALL-KU-32\33) show no persistence. In Table 1, DIF and SPEAR are near zero for these cases, and in Table 2, TCS is near zero, PRW is near one-half, and CP and CHI are near one. In small samples, the unconditional mean of CP exceeds one, even when alphas are i.i.d. All other cases require some or all of the evaluation-period alphas to make the survival cut, and in the SY samples, this biases the persistence measures upward.

The effect is sharpest in the look-ahead biased methodologies. For example, in Table 1, the PLA methodology creates a difference (DIF) between the top and bottom 3R1E decile portfolios of 0.73% per year in the SY-ALL case, and the FLA methodology creates a difference of 0.68%. The change in decile break-points that takes place going from the PLA to the FLA methodologies actually reduces the persistence bias slightly. The funds that disappear after the ranking period tend to be high-volatility funds so they tend to come out of the top decile more than the bottom. This moves the top break point down, reducing the volatility spread between deciles.

The incremental pruning in the SB sample only removes high-volatility funds, and thereby reduces the cross-sectional dispersion in volatility that causes the BGIR persistence effect in the first place. For example, in the SY-ALL case in Table 1, the 3R1E DIF falls to 0.19% per year. To see the effect that pruning has on volatility, note that in the SY-ALL cases, 5\6 and 5–7\8–10, in Table 2, SIGBAR is 1.84% in the SB samples, roughly half as large as in the corresponding KU samples.

4.1.2. Reversal effects of multiperiod criteria

A number of previous papers note that if survival depends on average performance across several periods, survivorship bias should induce spurious reversals in performance. Intuitively, ranking-period losers must subsequently do well in the evaluation period in order to survive, but winners have room to do poorly. More precisely, if the survival criterion is based on performance over a period of time longer than the evaluation period, then the critical level of evaluation-period performance is higher the lower is the ranking-period performance. Thus, conditioning on surviving beyond the evaluation period induces a monotonic, negative relation between ranking-period performance and evaluation-period performance.

The question is whether or not this effect dominates the BGIR persistence effect when cross-sectional heteroskedasticity is present. In general, the multiyear reversal effect more than offsets the BGIR persistence effect. In most cases in which part or all of the evaluation-period alphas are survivor-biased, the persistence measures are biased downward. Again, the cleanest examples of this are under the IND specification, now with the MY criterion. Table 1 shows that DIF is negative in all of the survivor-biased MY cases, and in most of these cases, SPEAR is negative, too. For example, the 3R1E DIF under the PLA methodology is -4.39% in the MY-ALL case. Similarly, most of the survivor-biased PRW and TCS measures indicate reversals in the MY panel of Table 2.

Table 1 shows that going from PLA to FLA also reduces the multiyear reversal effect. For instance, the 3R1E DIF under the FLA methodology is only -1.99% in the MY-ALL case. Funds that disappear during the evaluation period tend to be poor ranking-period performers, because of the multiyear survival criterion. Eliminating them from the ranking therefore raises the lower decile break points. This increases the average ranking-period performance in the lower deciles, which reduces the critical level of evaluation-period performance required for survival, again because of the multiyear criterion. The new ranking therefore reduces average evaluation-period performance in the lower deciles and consequently reduces the reversal.

The incremental survivor bias in the SB sample, caused by eliminating funds that disappear after the evaluation period, further reduces the reversal effect in the cases with three-year ranking periods. In the MY cases in Table 1, the 3R DIFs and SPEARs become less negative going from KU-FLA to SB. For example, the 3R1E DIF rises to -0.64% in the SB sample with ALL data available. Similarly, in Table 2, the MY 5–7\8–10 persistence measures show less reversal with SB than KU. We offer the following intuition for this reduction in reversals. Consider three consecutive periods: the ranking period, the three years after the ranking period, and the four years after that. Note that going from look-ahead biased cases to the SB sample eliminates funds that survive the first two periods but then disappear. Consider those that disappear in the third period. Intuitively, their performance pattern must be good-bad-bad or bad-good-bad; those with good performance in both of the first two periods tend to be too good to die in the third. Those with poor performance in both of the first two periods are already gone from the look-ahead biased samples in the first place. Therefore, the final cut made for the SB sample tends to eliminate reversers.

Table 1
Persistence measures for performance-ranked portfolios under the IND specification

Mean persistence measures in 2500 33-year samples of fund alphas. Each fund's annual alpha is normally distributed with mean zero. Fund alphas are independent, identically distributed across time, and cross-sectionally heteroskedastic. The initial number of funds in each sample is 213 and the annual growth rate in the number of funds is 5.8%. In the Z samples, funds never disappear. In the SY and MY samples, the fund attrition rate is 3.6% per year. The SY criterion cuts funds each year based on their past year's performance, while the MY criterion cuts funds based on their average performance over the previous five years. In samples labeled ALL, all fund returns are available, but in MSS samples, the last two months of disappearing fund returns are missing. Persistence tests employ strategies that sort funds into decile portfolios based on their performance over a preceding ranking period and evaluate average portfolio performance over a subsequent evaluation period. The label nRmE indicates that the ranking period is n years and the evaluation period is m years. In the KU-IAA samples, portfolios are rebalanced if funds disappear during the evaluation period. The KU-PLA and KU-FLA samples eliminate funds that do not have data available for n years after the ranking period. The KU-PLA samples first rank all funds and then eliminate the disappearing funds, while the KU-FLA samples eliminate disappearing funds before setting the decile breakpoints. The SB samples include only funds with data available until the end of the sample period. AV1 and AV10 are average annualized evaluation-period alphas of the bottom and top decile portfolios, in percent. DIF is AV10 minus AV1. TDIF is a t-statistic for DIF using the time series of portfolio differences to estimate the standard error. SPEAR is the Spearman rank correlation between ranking-period ranks and evaluation-period ranks

Samples with no attrition													
				Z									
				AV1	AV10	DIF	TDIF	SPEAR					
			1R1E	0.00	0.00	0.00	0.00	0.00					
			3R1E	– 0.01	0.00	0.01	0.03	– 0.01					
			3R3E	0.00	0.00	0.01	0.02	– 0.01					
Samples with attrition													
				SY					MY				
				AV1	AV10	DIF	TDIF	SPEAR	AV1	AV10	DIF	TDIF	SPEAR
ALL	Full sample												
	KU	IAA	1R1E	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	– 0.01	0.01
			3R1E	0.00	– 0.01	– 0.01	– 0.03	– 0.01	0.01	0.00	– 0.01	– 0.03	– 0.01
			3R3E	0.00	– 0.01	– 0.01	– 0.05	– 0.02	0.01	0.01	0.00	– 0.01	– 0.01

	Partial look-ahead biased sample												
	KU	PLA	1R1E[a]										
			3R1E	0.66	1.39	0.73	2.94	0.19	4.77	0.38	− 4.39	− 9.29	− 0.46
			3R3E	0.48	0.97	0.49	3.03	0.19	2.71	0.21	− 2.50	− 9.34	− 0.48
	Look-ahead biased sample												
	KU	FLA	1R1E	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.01
			3R1E	0.59	1.28	0.68	3.19	0.20	2.37	0.37	− 1.99	− 6.46	− 0.25
			3R3E	0.43	0.89	0.47	3.25	0.20	1.33	0.21	− 1.12	− 5.99	− 0.27
	Survivor-biased sample												
	SB		1R1E	0.28	0.39	0.11	0.68	0.06	1.28	0.73	− 0.55	− 2.41	− 0.06
			3R1E	0.24	0.43	0.19	1.12	0.13	1.43	0.79	− 0.64	− 2.55	− 0.01
			3R3E	0.25	0.44	0.19	1.58	0.12	1.06	0.96	− 0.09	− 0.65	0.10
MSS	Full sample												
	KU	IAA	1R1E	0.15	0.23	0.08	0.35	0.08	0.16	0.06	− 0.10	− 0.35	− 0.11
			3R1E	0.11	0.25	0.14	0.54	0.14	0.30	0.03	− 0.27	− 0.83	− 0.24
			3R3E	0.11	0.24	0.13	0.80	0.16	0.29	0.06	− 0.23	− 1.09	− 0.26
	Look-ahead biased sample												
	KU	FLA	1R1E	0.88	1.33	0.46	2.19	0.11	0.89	0.31	− 0.58	− 1.98	− 0.26
			3R1E	0.53	1.11	0.58	2.77	0.19	2.37	0.83	− 1.55	− 5.04	0.00
			3R3E	0.59	1.19	0.61	4.33	0.18	1.58	0.73	− 0.84	− 4.75	0.06
	Survivor-biased sample												
	SB		1R1E	0.28	0.39	0.11	0.70	0.07	1.30	0.74	− 0.56	− 2.50	− 0.06
			3R1E	0.22	0.38	0.16	0.97	0.12	1.36	0.78	− 0.57	− 2.37	0.00
			3R3E	0.24	0.41	0.17	1.45	0.12	1.02	0.97	− 0.05	− 0.40	0.11

[a] Same as KU-IAA.

Table 2
Persistence measures for contingency tables under the IND specification

Mean persistence measures in 2500 33-year samples of fund alphas. Each fund's annual alpha is normally distributed with mean zero. Fund alphas are independent, identically distributed across time, and cross-sectionally heteroskedastic. The initial number of funds in each sample is 213 and the annual growth rate in the number of funds is 5.8%. In the *Z* samples, funds never disappear. In the SY and MY samples, the fund attrition rate is 3.6% per year. The SY criterion cuts funds each year based on their past year's performance, while the MY criterion cuts funds based on their average performance over the previous five years. In samples labeled ALL, all fund returns are available, but in MSS samples, the last two months of disappearing fund returns are missing. Persistence tests examine performance over ranking-period\evaluation-period pairs. For example, 5–7\8–10 indicates that the ranking period consists of years 5–7 of the sample period and the evaluation period consists of years 8–10. With the KU samples, tests incorporate all funds with data available throughout the ranking and evaluation periods. The SB samples include only funds with data available until the end of the sample period. SIGBAR is the average volatility of the funds in the test sample, in percent. ABAR1 is the average annualized alpha of test funds during the ranking period, in percent, and ABAR2 is the average evaluation-period alpha. Contingency table tests define winners and losers each period relative to the median performance of test funds in that period. PRW is the fraction of ranking-period winners that win in the evaluation period. CP is the cross-product ratio and CHI is a chi-squared statistic with one degree of freedom, formed from the counts of repeat winners. TCS is the slope coefficient in a cross-sectional regression of evaluation-period performance on ranking-period performance

			Z						
			SIGBAR	ABAR1	ABAR2	PRW	CP	CHI	TCS
		5\6	4.70	0.00	−0.01	0.500	1.02	1.07	−0.09
		32\33	4.70	0.00	0.01	0.500	1.01	1.01	−0.01
		5–7\8–10	4.70	0.00	0.01	0.503	1.04	1.07	0.03
		28–30\31–33	4.70	0.00	0.00	0.500	1.01	1.03	−0.01

			SY							MY						
			SIGBAR	ABAR1	ABAR2	PRW	CP	CHI	TCS	SIGBAR	ABAR1	ABAR2	PRW	CP	CHI	TCS
ALL	KU	5\6	3.92	0.53	0.01	0.501	1.05	1.07	0.06	4.51	0.24	0.00	0.503	1.04	1.03	0.05
		32\33	3.09	0.42	0.00	0.500	1.01	1.09	−0.07	3.59	0.20	0.00	0.500	1.01	1.04	−0.07
		5–7\8–10	3.29	0.35	0.28	0.512	1.15	1.30	1.30	3.89	0.82	0.30	0.495	0.99	1.28	−1.26
		28–30\31–33	2.55	0.26	0.23	0.514	1.13	1.75	3.07	3.02	0.67	0.26	0.499	1.00	1.14	−2.09

SB		5\6 32\33[a]	1.84	0.03	0.03	0.503	1.07	1.06	0.06	1.78	0.13	0.16	0.498	1.05	1.07	– 0.29
		5–7\8–10 28–30\31–33[a]	1.84	0.03	0.04	0.501	1.06	1.06	0.12	1.78	0.15	0.19	0.500	1.07	1.10	0.10
MSS	KU	5\6	3.75	0.47	0.51	0.507	1.09	1.11	0.89	4.37	0.42	0.30	0.499	1.01	1.07	– 0.49
		32\33	2.94	0.37	0.40	0.506	1.06	1.27	2.13	3.47	0.34	0.22	0.498	0.99	1.07	– 0.89
		5–7\8–10	3.15	0.31	0.38	0.517	1.19	1.40	1.78	3.75	0.82	0.52	0.504	1.07	1.24	– 0.77
		28–30\31–33	2.43	0.23	0.30	0.518	1.16	2.08	4.03	2.89	0.67	0.42	0.509	1.09	1.35	– 0.70
SB		5\6 32\33[a]	1.81	0.03	0.03	0.503	1.07	1.06	0.04	1.74	0.12	0.14	0.498	1.05	1.07	– 0.27
		5–7\8–10 28–30\31–33[a]	1.81	0.03	0.04	0.502	1.06	1.06	0.11	1.74	0.14	0.17	0.500	1.07	1.11	0.07

[a]Same as KU.

4.1.3. Missing data and survivor bias

Missing data alone create a survivor bias. In the contingency table methodology, missing the final two months of data disqualifies the fund's entire evaluation-period alpha from the persistence test, creating a look-ahead bias. Yet even in the portfolio strategy tests that include all available alphas, KU-IAA, missing data introduce a bias. The IAA strategy reinvests money from funds that fail to make the cut into the average surviving fund. When the final two months of the disappearing funds are missing, the methodology effectively replaces the performance of those funds over those two months with the average performance of funds that did make the cut. This survivor bias results in spurious persistence in the SY samples, because of the BGIR effect, and spurious reversals in the MY samples, because of the multiperiod reversal effect. For example, in Table 1, under SY, missing data create a difference between the top and bottom 3R1E decile portfolios of 0.14% per year, even when all available alphas are included (IAA). Under MY, missing data create a difference between the top and bottom 3R1E decile portfolios of -0.27% per year in the IAA case.

4.1.4. Survivor biases in the presence of true persistence

Under the HET and MA1 specifications, the biases described above are still present. To facilitate comparison of the results for different generating processes, Table 3 focuses on the DIF variable and shows all cases. To separate survivor bias from true persistence, Table 4 shows the changes in the DIF variable from one case to the next.

Although the magnitudes of the survivor biases vary across specifications, the directions are almost always the same. The look-ahead bias and the missing data bias are always positive under the SY criterion and negative under the MY criterion. The ranking effect always offsets the partial look-ahead bias. With each survival criterion, the incremental survivor bias tends to have the same sign for HET and MA1 as it does for IND. For example, with the MY criteria, under all three alpha processes, the total look-ahead bias is negative, and for the 3R cases, the incremental survivor bias is positive.

4.1.5. Comparison with the results of Carhart (1997)

The performance-ranked portfolio methodology that we employ to test for persistence in simulated samples is comparable to the methodology that Carhart (1997) uses with real mutual fund data. This enables us to make some inferences about the true data-generating processes. First, the degree of persistence that Carhart (1997) demonstrates in his Table 8 is too great to be the result of survivor bias alone. In his full sample, he finds a difference of 0.24% per month between average four-factor alphas of top and bottom decile portfolios ranked on past one-year returns, and a difference of 0.36% per month between alphas of top and bottom decile portfolios ranked on past three-year alphas. In our full samples (KU-IAA) with no true persistence (IND), persistence levels never reach that magnitude. The cases with all data available show no persistence. In the cases with missing data, the SY samples show only a small degree of persistence (DIFs of 0.08% per year and 0.14% per year for the 1R1E and 3R1E samples, respectively). The MY full samples with missing data show only reversals.

Second, the pattern of look-ahead bias and incremental survivor bias that Carhart's results exhibit is different from the pattern we find. In Carhart's Table 8, look-ahead bias reduces persistence measures and survivor bias further reduces the persistence measures. For example, Carhart's 3R1E DIF falls from 0.36% per month in the full sample to 0.33% in the look-ahead biased sample and 0.18% in the survivor-biased sample. In other words, both the look-ahead bias and the incremental survivor bias are negative. By contrast, in all the cases we consider, the incremental survivor bias offsets the look-ahead bias. In particular, Table 4 shows that, while the look-ahead bias is always negative in the MY cases, the incremental survivor bias is almost always positive.

This comparison suggests that the true generating processes are quite different from the three that we consider. In particular, the pattern and magnitudes of DIF that Carhart reports cannot be explained without real persistence, even when the final returns of disappearing funds are missing. Thus, U.S. mutual fund performance is persistent, but the generating process is not captured by either of our persistence specifications.

4.2. Attrition effects

Both Brown and Goetzmann (1995) and Carhart (1997) provide evidence that survival criteria for real mutual funds are multiperiod. The simulated performance studies suggest that in this case, survivor bias tends to reduce measures of persistence, consistent with the findings of Hendricks et al. (1993) and Carhart (1997) that survivor-biased samples show less persistence than full samples. The fact that most mutual fund studies find persistence, despite some degree of survivor bias stemming from limited data sets, look-ahead methodologies, or missing returns, implies that performance truly persists. In particular, the previous subsection shows that the pattern and magnitudes of the DIF numbers reported by Carhart cannot be explained without real performance persistence.

Fully exploiting true persistence requires some knowledge of the true return-generating process. For example, under the HET specification, the longer the ranking period, the more informative is the ranking-period performance regarding which funds are best. Table 5, which gives persistence measures for both the HET and MA1 specifications, shows that under HET, the top decile performance, AV10, is better with three-year ranking periods than one-year. On the other hand, if fund performance is autocorrelated, the optimal ranking and holding periods depend on the autocorrelation structure. Table 5 shows that, under the MA1 specification, one-year ranking and holding periods work best.

Table 3
DIF persistence measures for various samples

Mean persistence measures in 2500 33-year samples of fund alphas. In IND samples, each fund's annual alpha is normally distributed with mean zero. Fund alphas are independent, identically distributed across time, and cross-sectionally heteroskedastic. In HET samples, mean fund returns vary cross-sectionally. In MA1 samples, fund returns follow mean zero MA (1) processes. The initial number of funds in each sample is 213 and the annual growth rate in the number of funds is 5.8%. In the Z samples, funds never disappear. In the SY and MY samples, the fund attrition rate is 3.6% per year. The SY criterion cuts funds each year based on their past year's performance, while the MY criterion cuts funds based on their average performance over the previous five years. In samples labeled ALL, all fund returns are available, but in MSS samples, the last two months of disappearing fund returns are missing. Persistence measures employ strategies that sort funds into decile portfolios based on their performance over a preceding ranking period and evaluate average portfolio performance over a subsequent evaluation period. The label nRmE indicates that the ranking period is n years and the evaluation period is m years. In the KU-IAA samples, portfolios are rebalanced if funds disappear during the evaluation period. The KU-PLA and KU-FLA samples eliminate funds that do not have data available for 4 years after the ranking period. The KU-PLA samples first rank all funds and then eliminate the disappearing funds, while the KU-FLA samples eliminate disappearing funds before setting the decile breakpoints. The SB samples include only funds with data available until the end of the sample period. DIF is the average difference between annualized evaluation-period alphas of the top and bottom decile portfolios, in per cent

Samples with no attrition

	Z		
	IND	HET	MA1
1RIE	0.00	1.75	1.76
3RIE	0.01	2.93	0.98
3R3E	0.01	2.93	0.32

Samples with attrition

			SY			MY		
			IND	HET	MA1	IND	HET	MA1
ALL	Full sample							
	KU IAA	1R1E	0.00	2.17	2.24	0.00	1.65	1.92
		3R1E	− 0.01	3.55	1.25	− 0.01	2.60	1.07
		3R3E	− 0.01	3.63	0.43	0.00	2.25	0.17
	Partial look-ahead biased sample							
	KU PLA	1R1E[a]						
		3R1E	0.73	4.04	1.87	− 4.39	− 3.04	− 3.77
		3R3E	0.49	3.92	0.88	− 2.50	− 0.98	− 2.56
	Look-ahead biased sample							
	KU FLA	1R1E	0.00	2.16	2.22	0.00	1.65	1.93
		3R1E	0.68	3.95	1.82	− 1.99	0.09	− 1.43
		3R3E	0.47	3.83	0.85	− 1.12	0.98	− 1.16
	Survivor-biased sample							
	SB	1R1E	0.11	2.64	2.85	− 0.55	0.76	1.39
		3R1E	0.19	3.91	1.57	− 0.64	1.64	− 0.03
		3R3E	0.19	3.90	0.67	− 0.09	2.08	− 0.15

Table 3 (*continued*)

				SY			MY		
				IND	HET	MA1	IND	HET	MA1
MSS	Full sample								
	KU	IAA	1R1E	0.08	2.22	2.26	− 0.10	1.53	1.81
			3R1E	0.14	3.57	1.34	− 0.27	2.26	0.76
			3R3E	0.13	3.64	0.54	− 0.23	1.98	− 0.08
	Look-ahead biased sample								
	KU	FLA	1R1E	0.46	2.46	2.46	− 0.58	0.97	1.24
			3R1E	0.58	3.94	1.76	− 1.55	0.37	− 1.10
			3R3E	0.61	3.93	1.01	− 0.84	1.04	− 1.04
	Survivor-biased sample								
	SB		1R1E	0.11	2.66	2.88	− 0.56	0.75	1.39
			3R1E	0.16	3.91	1.56	− 0.57	1.73	0.05
			3R3E	0.17	3.90	0.65	− 0.05	2.15	− 0.10

[a]Same as KU-IAA.

Researchers trying to estimate the parameters of the true return-generating process generally want a data set that is as free as possible from survivor bias. Even in the absence of survivor bias, however, fund attrition alters persistence measures because it alters the composition of funds in the sample. In order to draw correct inferences, the estimation procedure must account for the impact of fund attrition on sample moments.

To illustrate the pure attrition effect, we compare persistence measures for attrition-free samples, Z and ZS, with persistence measures for samples in which funds disappear but which are free of any form of survivor bias. The cases that are completely free of survivor bias but show attrition effects are the ALL-KU-IAA cases with the performance-ranked portfolios and the ALL-KU contingency tables and cross-sectional regressions with one-year ranking and evaluation periods.

Under the IND specification, the cases that are free of all forms of survivor bias show zero persistence, with or without attrition, so the attrition effect is zero. However, under the HET and MA1 specifications, attrition alters the persistence measures. Table 4 shows the incremental attrition effect in the DIF measure and Table 5 contains the complete set of persistence measures for samples with and without attrition.

In general, attrition prunes high-volatility funds from the sample, as the values of SIGBAR in Panel B of Table 5 show. This increases persistence measures under the HET specification because ranking on past performance more precisely ranks on true mean when volatility is low. However, under the

Table 4
Incremental attrition effects and survivor biases in the DIF persistence measure

Persistence measures are simulated from 2500 33-year samples of fund alphas. In IND samples, each fund's annual alpha is normally distributed with mean zero. Fund alphas are independent, identically distributed across time, and cross-sectionally heteroskedastic. In HET samples, mean fund returns vary cross-sectionally. In MA1 samples, fund returns follow mean zero MA(1) processes. The initial number of funds in each sample is 213 and the annual growth rate in the number of funds is 5.8%. In the *Z* samples, funds never disappear. In the SY and MY samples, the fund attrition rate is 3.6% per year. The SY criterion cuts funds each year based on their past year's performance, while the MY criterion cuts funds based on their average performance over the previous five years. In samples labeled ALL, all fund returns are available, but in MSS samples, the last two months of disappearing fund returns are missing. Persistence measures employ strategies that sort funds into decile portfolios based on their performance over a preceding ranking period and evaluate average portfolio performance over a subsequent evaluation period. The label nRmE indicates that the ranking period is n years and the evaluation period is m years. In the KU-IAA samples, portfolios are rebalanced if funds disappear during the evaluation period. The KU-PLA and KU-FLA samples eliminate funds that do not have data available for n years after the ranking period. The KU-PLA samples first rank all funds and then eliminate the disappearing funds, while the KU-FLA samples eliminate disappearing funds before setting the decile breakpoints. The SB samples include only funds with data available until the end of the sample period. DIF is the average difference between annualized evaluation-period alphas of the bottom and top decile portfolios, in percent. The table shows differences in the DIF measures between various cases

				SY			MY		
				IND	HET	MA1	IND	HET	MA1
Attrition effect	ALL	IAA minus Z	1R1E	0.00	0.42	0.48	0.00	– 0.09	0.17
			3R1E	– 0.02	0.61	0.27	– 0.02	– 0.33	0.09
			3R3E	– 0.02	0.70	0.11	– 0.01	– 0.68	– 0.15
Partial look-ahead bias	ALL	PLA minus IAA	1R1E	0.00	0.00	0.00	0.00	0.00	0.00
			3R1E	0.73	0.49	0.63	– 4.39	– 5.64	– 4.84
			3R3E	0.50	0.29	0.45	– 2.50	– 3.23	– 2.73
Ranking effect	ALL	FLA minus PLA	1R1E	0.00	– 0.01	– 0.02	0.00	0.00	0.01
			3R1E	– 0.04	– 0.09	– 0.05	2.40	3.13	2.34
			3R3E	– 0.03	– 0.09	– 0.02	1.38	1.96	1.40

Incremental survivor bias	ALL	SB minus FLA	1R1E	0.11	0.47	0.63	− 0.55	− 0.89	− 0.55
			3R1E	− 0.49	− 0.04	− 0.25	1.35	1.55	1.40
			3R3E	− 0.28	0.07	− 0.19	1.03	1.10	1.00
Missing data bias	MSS minus ALL	IAA	1R1E	0.08	0.04	0.02	− 0.10	− 0.12	− 0.11
			3R1E	0.14	0.02	0.09	− 0.26	− 0.34	− 0.31
			3R3E	0.14	0.01	0.11	− 0.23	− 0.27	− 0.25
Look-ahead bias	MSS	FLA minus IAA	1R1E	0.38	0.25	0.19	− 0.47	− 0.56	− 0.57
			3R1E	0.44	0.37	0.43	− 1.27	− 1.89	− 1.86
			3R3E	0.47	0.29	0.46	− 0.61	− 0.93	− 0.96
Incremental survivor bias	MSS	SB minus FLA	1R1E	− 0.35	0.20	0.42	0.02	− 0.22	0.16
			3R1E	− 0.42	− 0.03	− 0.21	0.97	1.36	1.15
			3R3E	− 0.44	− 0.03	− 0.35	0.79	1.10	0.94

Table 5
Attrition effects in the presence of true performance persistence

Mean persistence measures in 2500 33-year samples of fund alphas. Fund alphas are cross-sectionally independent and heteroskedastic. In HET samples, mean fund returns vary cross-sectionally. In MA1 samples, fund returns follow mean zero MA(1) processes. The initial number of funds in each sample is 213 and the annual growth rate in the number of funds is 5.8%. In the *Z* samples, funds never disappear. In the SY and MY samples, the fund attrition rate is 3.6% per year. The SY criterion cuts funds each year based on their past year's performance, while the MY criterion cuts funds based on their average performance over the previous five years. Persistence tests in Panel A employ strategies that sort funds into decile portfolios based on their performance over a preceding ranking period and evaluate average portfolio performance over a subsequent evaluation period. When the SY or MY criterion is applied, all fund returns are available, portfolios are rebalanced if funds disappear during the evaluation period, and the sample is labeled ALL-KU-IAA. The label nRmE indicates that the ranking period is n years and the evaluation period is m years. AV1 and AV10 are average annualized evaluation-period alphas of the bottom and top decile portfolios, in percent. DIF is AV10 minus AV1. TDIF is a t-statistic for DIF using the time series of portfolio differences to estimate the standard error. SPEAR is the Spearman rank correlation between ranking-period ranks and evaluation-period ranks. Persistence tests in Panel B examine performance over ranking-period\evaluation-period pairs. For example, 5\6 indicates that the ranking period consists of year 5 of the sample period and the evaluation period consists of year 6. When the SY or MY criterion is applied, the sample is labeled ALL-KU, and includes all funds with data available throughout the ranking and evaluation periods. SIGBAR is the average volatility of the funds in the test sample, in percent. ABAR1 is the average annualized alpha of test funds during the ranking period, in percent, and ABAR2 is the average evaluation-period alpha. Contingency table tests define winners and losers each period relative to the median performance of test funds in that period. PRW is the fraction of ranking-period winners that win in the evaluation period. CP is the cross-product ratio and CHI is a chi-squared statistic with one degree of freedom, formed from the counts of repeat winners. TCS is the slope coefficient in a cross-sectional regression of evaluation-period performance on ranking-period performance

Panel A: Persistence measures for performance-ranked portfolio strategies

		HET					MA1				
		AV1	AV10	DIF	TDIF	SPEAR	AV1	AV10	DIF	TDIF	SPEAR
ZS	1R1E	−0.88	0.88	1.75	3.31	0.93	−0.88	0.87	1.75	3.29	0.93
	3R1E	−1.47	1.45	2.92	5.11	0.97	−0.50	0.48	0.99	1.68	0.75
	3R3E	−1.46	1.45	2.91	7.69	0.98	−0.17	0.16	0.33	0.83	0.43
Z	1R1E	−0.87	0.87	1.75	5.39	0.96	−0.88	0.88	1.76	5.39	0.97
	3R1E	−1.46	1.47	2.93	8.27	0.99	−0.50	0.49	0.98	2.73	0.90
	3R3E	−1.46	1.47	2.93	12.60	1.00	−0.17	0.16	0.32	1.37	0.64

SY	ALL KU	IAA	1R1E	− 0.77	1.40	2.17	7.59	0.99			− 1.13	1.11	2.24	7.82	0.99		
			3R1E	− 1.54	2.00	3.55	13.22	1.00			− 0.63	0.62	1.25	4.50	0.95		
			3R3E	− 1.52	2.11	3.63	20.19	1.00			− 0.16	0.27	0.43	2.33	0.78		
MY	ALL KU	IAA	1R1E	− 0.34	1.32	1.65	5.18	0.95			− 0.95	0.97	1.92	6.02	0.98		
			3R1E	− 0.73	1.87	2.60	7.02	0.99			− 0.54	0.53	1.07	2.90	0.92		
			3R3E	− 0.37	1.88	2.25	7.95	0.97			0.01	0.18	0.17	0.63	0.49		

Panel B: Persistence measures for contingency tables and cross-sectional regressions

			HET							MA1						
			SIGBAR	ABAR1	ABAR2	PRW	CP	CHI	TCS	SIGBAR	ABAR1	ABAR2	PRW	CP	CHI	TCS
ZS		5\6	4.72	− 0.02	− 0.02	0.58	2.06	3.43	0.92	4.71	0.00	0.01	0.58	2.09	3.54	0.91
		32\33	4.71	0.00	− 0.01	0.58	1.97	13.13	1.85	4.71	0.00	− 0.01	0.58	1.96	12.96	1.84
Z		5\6	4.71	0.01	0.00	0.58	1.99	8.00	1.48	4.70	0.00	0.00	0.58	1.99	7.96	1.48
		32\33	4.70	0.01	0.00	0.58	1.95	33.77	3.02	4.70	0.00	0.00	0.58	1.95	33.59	3.04
SY	ALL KU	5\6	3.96	0.67	0.13	0.59	2.16	9.36	1.82	3.95	0.60	0.04	0.59	2.13	9.16	1.87
		32\33	3.16	0.71	0.29	0.61	2.52	60.25	4.97	3.15	0.49	0.05	0.61	2.42	55.44	4.94
MY	ALL KU	5\6	4.55	0.29	0.07	0.58	1.98	7.58	1.34	4.53	0.27	0.01	0.58	2.01	7.94	1.49
		32\33	3.78	0.68	0.51	0.60	2.31	50.14	3.27	3.74	0.25	0.03	0.60	2.26	47.49	3.68

HET specification, attrition also prunes funds with low means. This reduces persistence measures by reducing the cross-sectional dispersion in mean performance. The net effect depends on the survival criterion. Table 4 shows that with the SY criterion, the first effect dominates, whereas the second effect dominates with the MY criterion. For example, under HET, attrition increases the 3R1E DIF by 0.61% under the SY criterion, but reduces it by 0.33% under the MY criterion. By cutting based on five-year performance rather than one-year performance, the MY criterion cuts relatively more low-mean funds. This is because the cross-sectional dispersion of means in average five-year returns is the same as in one-year returns but the dispersion of volatilities is smaller.

Under the MA1 specification, the attrition effect is positive in general. From Eqs. (10) and (11), the MA coefficient θ_i is decreasing in σ_i. Culling high-volatility funds increases the average MA coefficient in the sample, which increases measures of persistence. In Table 4, the attrition effect is positive for all MA1 except MY-3R3E.

With the MY criterion and the MA1 alpha process, rebalancing the decile portfolio because of fund attrition creates a reversal effect. Because we generate alphas annually, this result of rolling dying funds into survivors is only apparent in the 3R3E case. To understand the effect, first consider the funds that are alive later in the evaluation period. They must have done well enough early in the evaluation period to endure. Having done well early, they are likely to do well later in the evaluation period because of the moving average process. Therefore, fund attrition increases each decile's average performance late in the evaluation period. Next, note that with the MY criterion, the increase is greater for the lower deciles than for the higher deciles. This is because, under the MY criterion, the funds in the lower deciles have higher survival hurdles. Given that they survive early in the evaluation period, their early performance must be better than that of funds that survived early in higher deciles. This makes their performance later in the evaluation period better than that of higher-decile funds. Rolling dying funds into survivors therefore gives the lower-decile portfolios more of a boost and this creates a tendency for reversal. For example, in Table 5, AV1, the average bottom-decile performance, is actually positive for the MY-ALL-KU-IAA-3R3E case, though it is negative for Z-3R3E and SY-ALL-KU-IAA-3R3E.

4.3. Test specification and power

Tables 6 and 7 provide information about the specification and power of various persistence tests. Table 6 examines test specification with upper and lower rejection frequencies of two-sided 5% tests in samples with no true persistence. Table 7 demonstrates the power of one-sided 2.5% tests for persistence in samples with true persistence. The tables also contain rejection rates for the HPZJ test which we present as two one-sided 2.5% tests, one for true persistence and one for spurious persistence.

Overall, in the absence of survivor bias, the DIF t-test using one-year evaluation periods appears to be the best specified under the null hypothesis of no persistence and one of the most powerful against the alternatives that we consider. Also well-specified and powerful in large samples, the chi-squared test is the most robust to the presence of survivor bias.

4.3.1. Test specification

Consider the rejection rates of the persistence tests under the IND specification in Table 6 for the survivor-bias-free cases. The survivor-bias-free cases are the Z cases, the KU-ALL-IAA decile tests, and the KU-ALL contingency tables with one-year evaluation periods. The upper and lower rejection rates for the TDIF test in Table 6 are close to 2.5% for both the SY and MY criteria when the evaluation period is one year. With the 3R3E case, DIF overrejects because the Newey–West standard errors do not fully correct for overlapping evaluation periods in small samples.

The Spearman rank test overrejects slightly, with rates of about 9% in a 5% test. With cross-sectional heteroskedasticity, the same funds consistently show extreme performance, and this can create the illusion of persistent performance.

The chi-squared tests are reasonably well specified. However, the cross-sectional t-tests substantially overreject the null hypothesis because the alphas are heteroskedastic. In particular, OLS underestimates the volatility of the slope coefficient because extreme values of α_{t-1} are associated with large volatilities of α_t. While the direction of the effect is not surprising, the magnitudes are remarkable. Rejection rates are 25–30% in a 5% test. In the survivor-bias-free cases, the HPZJ tests for true and spurious persistence also overreject with frequencies of 5–7% in 2.5% tests.

Table 6 also shows how the tests fare in the presence of two forms of survivor bias: MSS, the bias introduced when the last two months of disappearing fund returns are missing, and SB, the sample selection method that includes only funds with data available until the end of the sample period. Most recent mutual fund studies incorporate all available data, so they are only subject to the missing data bias. However, new hedge fund data sets can suffer from more severe survivorship bias (see, for example, Ackermann et al., 1996; Brown et al., 1997).

The chi-squared tests with one-year ranking and evaluation periods are the most robust to the presence of survivorship bias. The directional bias in these tests is slight and the total rejection frequencies remain between 5% and 8% in 5% tests. The Spearman test is relatively robust to the milder missing data bias, but it underrejects in the SB samples. For example, under the MY survival criterion, the Spearman rejection rates are zero. The DIF t-test is also relatively robust in MSS samples, but is severely biased in SB samples. Under the MY

Table 6

Specification of persistence tests: Rejection frequencies under the IND process

Upper and lower rejection frequencies at the indicated significance level in per cent, for various tests in 2500 33-year samples of fund alphas. For all but the HPZJ statistics, the table indicates rejection rates in two-sided tests for persistence. For HPZJ-true, the table indicates rejection rates for one-sided 2.5% tests for true persistence, and for HPZJ-spur, the table indicates rejection rates for one-sided 2.5% tests for a J-shape in the past performance–present performance relation. Each fund's annual alpha is normally distributed with mean zero. Fund alphas are independent, identically distributed across time, and cross-sectionally heteroskedastic. Fund alphas are cross-sectionally independent and heteroskedastic. The initial number of funds in each sample is 213 and the annual growth rate in the number of funds is 5.8%. In the Z samples, funds never disappear. In the SY and MY samples, the fund attrition rate is 3.6% per year. The SY criterion cuts funds each year based on their past year's performance, while the MY criterion cuts funds based on their average performance over the previous five years. In samples labeled ALL, all fund returns are available, but in MSS samples, the last two months of disappearing fund returns are missing. The KU samples keep funds until their data become unavailable. The SB samples include only funds with data available until the end of the sample period. N is the average number of funds in the test sample per year. The TDIF, SPEAR, and HPZJ tests employ strategies that sort funds into decile portfolios based on their performance over a preceding ranking period and evaluate average portfolio performance over a subsequent evaluation period. The label nRmE indicates that the ranking period is n years and the evaluation period is m years. With the KU samples, portfolios are rebalanced if fund data become unavailable during the evaluation period. TDIF is a t-statistic for the difference in top and bottom decile portfolios using the time series of portfolio differences to estimate the standard error. SPEAR is the Spearman rank correlation between ranking-period ranks (RPR) and evaluation-period ranks (EPR). HPZJ is the t-statistic for the linear term in the regression of EPR on a constant, RPR, and RPR^2. The CHI and TCS persistence tests examine performance over ranking-period\evaluation-period pairs. For example, 5–7\8–10 indicates that the ranking period consists of years 5–7 of the sample period and the evaluation period consists of years 8–10. CHI is a chi-squared statistic with one degree of freedom, formed from the counts of repeat winners. TCS is the slope coefficient in a cross–sectional regression of evaluation-period performance on ranking-period performance. With the KU samples, the CHI and TCS tests incorporate all funds with data available throughout the ranking and evaluation periods

	Aver. N per year		Z		KU ALL				KU MSS				SB MSS			
					SY		MY		SY		MY		SY		MY	
	ZS	Z	<2.5%	>97.5%	<2.5%	>97.5%	<2.5%	>97.5%	<2.5%	>97.5%	<2.5%	>97.5%	<2.5%	>97.5%	<2.5%	>97.5%
TDIF																
1R1E	236	630	2.7	2.4	2.4	2.6	2.4	2.5	1.2	5.4	4.9	1.1	0.3	9.8	63.9	0.0
3R1E	205	548	2.4	2.4	2.6	2.6	2.6	2.4	0.8	7.8	11.4	0.5	0.2	14.4	58.6	0.0
3R3E	205	548	4.9	5.5	6.2	4.6	4.5	4.0	1.7	16.5	19.9	0.4	0.2	28.8	12.0	4.2

SPEAR																
1R1E	236	630	4.6	4.2	4.8	3.6	3.9	5.4	1.6	3.6	7.0	2.2	0.1	0.5	0.0	0.0
3R1E	205	548	4.3	4.0	4.8	3.9	4.3	4.4	1.4	6.0	11.0	0.6	0.0	1.8	0.0	0.0
3R3E	205	548	7.2	6.1	7.3	6.1	6.0	5.6	1.7	7.3	13.2	0.6	0.0	0.6	0.0	0.0
CHI																
5\6	99	267	3.3	2.5	3.4	3.7	1.7	3.8	1.6	4.2	3.0	3.1	2.7	2.8	3.5	2.1
32\33[a]	456	1218	2.1	2.7	3.0	3.2	2.7	2.5	1.0	6.7	4.0	1.7				
5–7\8–10	99	267	3.3	2.4	1.1	6.2	5.9	2.8	1.0	8.5	3.5	4.8	2.6	2.3	3.6	2.6
28–30\31–33	364	972	2.0	3.0	0.6	14.7	3.5	3.2	0.4	17.7	1.2	8.4				
TCS																
5\6	99	267	14.4	12.3	11.4	12.4	13.2	14.1	3.4	25.1	20.9	8.3	8.8	9.4	20.4	14.1
32\33[a]	456	1218	12.2	13.2	16.4	15.1	16.7	15.2	1.3	53.1	28.8	7.3				
5–7\8–10	99	267	12.0	12.2	2.6	33.3	35.5	5.3	1.0	44.3	27.8	8.8	8.8	11.5	16.7	17.5
28–30\31–33[a]	364	972	13.7	13.2	0.5	69.8	52.6	3.6	0.1	85.0	29.2	11.8				
HPJZ-true[b]																
1R1E	236	630		5.9		7.5		5.5		0.4		1.2		0.0		0.0
3R1E	205	548		6.0		6.0		6.2		0.6		0.5		0.0		0.0
3R3E	205	548		8.2		8.1		9.4		0.6		0.6		0.0		0.0
Specification and power of HPZ test for spurious persistence																
HPJZ-spur[b]																
1R1E	236	630	5.8		6.0		6.3		29.3		16.4		81.5		99.9	
3R1E	205	548	5.6		4.3		6.8		23.0		21.0		64.1		99.5	
3R3E	205	548	7.9		7.0		9.9		39.2		34.6		90.3		100.0	

[a] SB-MSS same as KU-MSS.
[b] One-sided test.

Table 7
Power of persistence tests: Rejection frequencies under the HET and MA1 processes

Rejection frequencies in per cent, for various tests in 2500 33-year samples of fund alphas. For all statistics except HPZJ-spur, the table indicates rejection rates in one-sided 2.5% tests for true persistence. For HPZJ-spur, the table indicates rejection rates in one-sided 2.5% tests for a J-shape in the past performance–present performance relation. Fund alphas are cross-sectionally independent and heteroskedastic. Under HET, fund alphas are identically distributed through time, but mean alphas vary across funds. Under MA1, each fund's alpha follows a zero-mean moving average process of order one. The initial number of funds in each sample is 213 and the annual growth rate in the number of funds is 5.8%. In the Z samples, funds never disappear. In the SY and MY samples, the fund attrition rate is 3.6% per year. The SY criterion cuts funds each year based on their past year's performance, while the MY criterion cuts funds based on their average performance over the previous five years. In samples labeled ALL, all fund returns are available, but in MSS samples, the last two months of disappearing fund returns are missing. The KU samples keep funds until their data become unavailable. The SB samples include only funds with data available until the end of the sample period. The TDIF, SPEAR, and HPZJ tests employ strategies that sort funds into decile portfolios based on their performance over a preceding ranking period and evaluate average portfolio performance over a subsequent evaluation period. The label nRmE indicates that the ranking period is n years and the evaluation period is m years. With the KU samples, portfolios are rebalanced if fund data become unavailable during the evaluation period. TDIF is a t-statistic for the difference in top and bottom decile portfolios using the time series of portfolio differences to estimate the standard error. SPEAR is the Spearman rank correlation between ranking-period ranks (RPR) and evaluation-period ranks (EPR). HPZJ is the t-statistic for the linear term in the regression of EPR on a constant, RPR, and RPR^2. The CHI and TCS persistence tests examine performance over ranking-period\evaluation-period pairs. For example, 5–7\8–10 indicates that the ranking period consists of years 5–7 of the sample period and the evaluation period consists of years 8–10. CHI is a chi-squared statistic with one degree of freedom, formed from the counts of repeat winners. TCS is the slope coefficient in a cross-sectional regression of evaluation-period performance on ranking-period performance. With the KU samples, the CHI and TCS tests incorporate all funds with data available throughout the ranking and evaluation periods

	HET								MA1						
	ZS (%)	Z (%)	KU ALL		KU MSS		SB MSS		ZS (%)	Z (%)	KU ALL		KU MSS		SB MSS
			SY (%)	MY (%)	SY (%)	MY (%)	SY (%)	MY (%)			SY (%)	MY (%)	SY (%)	MY (%)	SY (%)
TDIF															
1R1E	85	100	100	100	100	99	100	71	86	100	100	100	100	100	100
3R1E	99	100	100	100	100	100	100	100	35	71	98	76	100	56	100
3R3E	100	100	100	100	100	100	100	100	16	28	55	10	76	2	95

SPEAR															
1R1E	100	100	100	100	100	100	100	10	100	100	100	100	100	100	100
3R1E	100	100	100	100	100	100	100	98	79	99	100	99	100	96	100
3R3E	100	100	100	100	100	100	100	100	34	61	82	37	86	8	62
CHI															
5\6	30	73	84	74	85	71	89	81	31	74	83	76	83	74	88
32\33[a]	93	100	100	100	100	100			93	100	100	100	100	100	
5–7\8–10	78	100	100	99	100	98	100	100	4	8	17	6	21	7	11
28–30\31–33[a]	100	100	100	100	100	100			9	20	62	14	68	23	
TCS															
5\6	27	38	47	34	68	23	84	44	25	37	48	38	68	28	90
32\33[a]	47	73	95	75	100	61			47	73	95	83	100	65	
5–7\8–10	59	87	99	47	100	51	100	92	13	16	39	4	50	5	23
28–30\31–33[a]	93	100	100	93	100	97			16	24	85	2	94	5	
HPJZ-true															
1R1E	46	72	88	50	89	37	62	0	49	72	97	82	98	78	100
3R1E	70	95	100	86	100	68	100	0	14	31	64	45	59	21	58
3R3E	90	100	100	61	100	29	100	0	13	16	16	8	8	1	2
HPJZ-spur															
1R1E	0	0	0	0	0	0	0	99	0	0	0	0	0	0	0
3R1E	0	0	0	0	0	0	0	18	1	0	0	0	0	0	0
3R3E	0	0	0	0	0	0	0	2	5	2	2	10	5	32	31

[a] SB-MSS same as KU-MSS.

criterion, the DIF t-tests with one-year evaluation periods reject in favor of reversals about 60% of the time.

The HPZJ tests perform reasonably well in the survivor-biased cases. In Table 6, the HPZJ tests diagnose spurious persistence 20–100% of the time and rarely indicate true persistence. Brown et al. (1997) also simulate the HPZJ test and their results for the case of cross-sectionally independent fund returns are consistent with ours. However, they find that the test overrejects in favor of true persistence when fund returns are cross-sectionally correlated.

4.3.2. Power

Table 7 contains rejection rates for one-sided 2.5% tests for persistence under the HET and MA1 specifications. The degree of persistence introduced in these alternative hypotheses is conservative compared to that found in mutual fund data by Carhart (1997) and others. However, it is still so strong that many of the tests show 100% power. Nevertheless, the table provides some indication of the relative strengths of the different tests.

To understand the patterns in Table 7, note that, by construction, the HET and MA1 alpha processes produce the same persistence measures with one-year ranking and evaluation periods in the samples with no attrition. For example, for both the ZS and Z samples in Table 5, the 1R1E persistence measures are the same for the HET process as they are for the MA1 process. Similarly, the 5\6 and 32\33 persistence measures are the same for the HET process as they are for the MA1 process. However, with longer ranking and evaluation periods, the persistence measures under HET become stronger, while the persistence measures under MA1 become weaker. These features are apparent in Table 7. For both Z samples, the tests with one-year ranking and evaluation periods are equally powerful against the HET and MA1 alternative hypothesis. Longer ranking and evaluation periods make the persistence tests more powerful under the HET specification and less powerful under the MA1 specification.

Against the HET alternative, the Spearman test appears to be the most powerful in the absence of survivor bias, although both the TDIF and the chi-squared tests are 100% powerful in large samples with three-year ranking and evaluation periods. In particular, the chi-squared tests perform well at the end of the sample period when the number of funds is large.

Even with missing data, all three tests remain extremely powerful, even with one-year ranking and evaluation periods. However, in the SB sample under the MY criterion, the power of the 1R1E Spearman test drops to 10% while that of the TDIF test remains at 71%. Again, the chi-squared test is most robust to the survivor bias, with power ranging from 81% to 100%.

The results with the MA1 specification are similar. The Spearman test is the most powerful in the absence of survivor bias. The chi-squared test is powerful in the larger samples with one-year ranking and evaluation periods and is the most robust to the presence of survivor bias. Under both the HET and MA1 specifications, the cross-sectional t-test is generally the weakest.

Table 7 also illustrates the performance of the HPZJ tests in the presence of true persistence. In most cases, the tests for true persistence are fairly powerful and the tests for spurious persistence find none. However, in the SB sample with the MY criterion, the tests never indicate true persistence, only spurious persistence.

5. Conclusion

We simulate standard tests of performance persistence under a variety of assumptions about return-generating processes, survival criteria, and data availability. When survival depends on performance over several periods, survivorship bias induces spurious reversals, despite the presence of cross-sectional heteroskedasticity in performance. In light of evidence that survival criteria in the mutual fund industry are multiperiod, these results reinforce the conclusions of most empirical studies that fund performance is truly persistent. Similarly, our results suggest that the delisting biases in the CRSP data documented by Shumway (1997) cause empirical work to understate the level of persistence in stock returns.

When returns are truly persistent, the simulations also reveal an attrition effect, distinct from survivorship bias. Mean persistence measures in the sample with attrition differ from those in a hypothetical sample in which funds never disappear. The direction of the change depends on the precise nature of the return-generating process. The results suggest that researchers trying to estimate the parameters of the true process need to account for the impact of attrition on sample moments.

Finally, we examine the specification and power of the persistence tests. Overall, in the absence of survivor bias, the t-test for the difference between the top and bottom performance deciles appears to be the best specified under the null hypothesis of no persistence and the among the most powerful against alternatives. On the other hand, the chi-squared test performs the best in the presence of survivor bias.

References

Ackermann, C., McEnally, R., Ravenscraft, D., 1996. The performance of hedge funds: risk, return, and incentives. Journal of Finance 54, 833–874.

Brown, S., Goetzmann, W., 1995. Attrition and mutual fund performance. Journal of Finance 50, 679–698.

Brown, S., Goetzmann, W., Ibbotson, R., 1999. Offshore hedge funds: survival and performance, 1989–95. The Journal of Business 72, 91–117.

Brown, S., Goetzmann, W., Ibbotson, R., Ross, S., 1992. Survivorship bias in performance studies. Review of Financial Studies 5, 553–580.
Brown, S., Goetzmann, W., Ibbotson, R., Ross, S., 1997. Rejoinder: the J-shape of performance persistence given survivorship bias. The Review of Economics and Statistics 79, 161–166.
Carhart, M., 1997. Survivorship bias and mutual fund performance. Unpublished Working Paper. University of Southern California.
Carhart, M., 1998. On persistence in mutual fund performance. Journal of Finance 52, 57–82.
DeBondt, W., Thaler, R., 1985. Does the stock market overreact. Journal of Finance 40, 793–805.
Elton, E., Gruber, M., Blake, C., 1996a. The persistence of risk-adjusted mutual fund performance. Journal of Business 69, 133–157.
Elton, E., Gruber, M., Blake, C., 1996b. Survivorship bias and mutual fund performance. Review of Financial Studies 9, 1097–1120.
Fama, E., French, K., 1996. Multifactor explanations of asset pricing anomalies. Journal of Finance 51, 55–84.
Goetzmann, W., Ibbotson, R., 1994. Do winners repeat? Patterns in mutual fund performance. Journal of Portfolio Management 20, 9–17.
Grinblatt, M., Titman, S., 1992. The persistence of mutual fund performance. Journal of Finance 47, 1977–1984.
Gruber, M., 1996. Another puzzle: the growth in actively managed mutual funds. Journal of Finance 51, 783–810.
Hendricks, D., Patel, J., Zeckhauser, R., 1993. Hot hands in mutual funds: short-run persistence of relative performance, 1974–1988. Journal of Finance 48, 93–130.
Hendricks, D., Patel, J., Zeckhauser, R., 1997. The J-shape of performance persistence given survivorship bias. The Review of Economics and Statistics 79, 161–166.
Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 48, 65–91.
Kothari, S., Warner, J., 1997. Evaluating mutual fund performance. Unpublished Working Paper. University of Rochester.
Malkiel, B., 1995. Returns from investing in equity mutual funds 1971–1991. Journal of Finance 50, 549–572.
Shumway, T., 1997. The delisting bias in CRSP data. Journal of Finance 52, 327–340.
Wermers, R., 1997. Momentum investment strategies of mutual funds, performance persistence, and survivor bias. Unpublished Working Paper. University of Colorado.