Survivorship Bias in Performance Studies

Stephen J. Brown
New York University

William Goetzmann
Columbia University

Roger G. Ibbotson
Stephen A. Ross
Yale University

Recent evidence suggests that past mutual fund performance predicts future performance. We analyze the relationship between volatility and returns in a sample that is truncated by survivorship and show that this relationship gives rise to the appearance of predictability. We present some numerical examples to show that this effect can be strong enough to account for the strength of the evidence favoring return predictability.

Past performance does not guarantee future performance. Empirical work from the classic study by Cowles (1933) to work by Jensen (1968) suggests that there is only very limited evidence that professional money managers can outperform the market averages

The first-named author acknowledges support of a Yamaichi Faculty Fellowship. We thank, for their unusually constructive comments and support, Campbell Harvey, Thomas Philips, Richard Roll, the editor (Chester Spatt), the referee (Peter Bossaerts), participants in presentations at Berkeley, Columbia, Cornell, New York University, Stanford, University of Massachusetts at Amherst, Vanderbilt University, Washington University at St. Louis, the 1991 Johnson Symposium at the University of Wisconsin, the Second Conference on Finance and Accounting at the University of Buffalo, 1991, and the 1992 Western Finance Association meetings. Remaining errors are our own. Address correspondence to Stephen J. Brown, Department of Finance, Stern School of Business, New York University, 44 West 4th St., New York, NY 10012-1126.

The Review of Financial Studies 1992 Volume 5, number 4, pp. 553–580

on a risk-adjusted basis. While more recent evidence[1] qualifies this negative conclusion somewhat [Grinblatt and Titman (1989), Ippolito (1988)], there is still no strong evidence that manager performance over and above the market indices can justify the fees managers charge and the commission costs they incur.

The fact that managers as a group perform poorly does not preclude the possibility that particular managers have special skills. Given the high turnover of managers, it is conceivable that the market selects out those managers with skills. Skillful managers are those who succeed and survive. It is this view, fostered by annual mutual fund performance reviews of the type published by *Barrons, Business Week, Consumer Reports,* and other publications, that leads to the popular investment strategy of selling shares in mutual funds that underperform the average manager in any given year, and buying shares in those funds with superior performance. Despite the popular impression that "hot hands" exist among mutual funds, there has been very limited empirical evidence to address this issue.

Past performance is usually a highly significant input into the decision to hire or fire pension fund money managers. However, Kritzman (1983) reports that for fixed-income pension fund money managers retained for at least 10 years, there is no relationship either in returns or in relative rankings between the performance in the first five years and the second five years. In an unpublished portion of the same study, this finding also extended to equity managers. Similar results are found for institutional funds by Dunn and Theisen (1983) and for commodity funds by Elton, Gruber, and Rentzler (1990).[2] In contrast to these findings, Elton and Gruber (1989, p. 602) conclude on the basis of a Securities and Exchange Commission (1971) study that mutual funds which outperform other funds in one period will tend to outperform them in a second. Grinblatt and Titman (1988) suggest that five-year risk-adjusted mutual fund returns do contain some predictive power for subsequent returns. Lehmann and Modest report similar results for the period 1968–1982, but suggest that this finding is sensitive to the method used to compute risk-adjusted performance measures.

On the basis of data for the period 1974–1988 both Hendricks, Patel, and Zeckhauser (1991) and Goetzmann and Ibbotson (1991) obtain far stronger results. The first study is limited to 165 equity

[1] Some of this evidence is controversial in nature. See Elton et al. (1993) for a discussion of the Ippolito findings.

[2] The commodity fund result applies to returns on funds. However, Elton, Gruber, and Rentzler (1990) find evidence of persistence in performance of different funds managed by the same general partner. It would be interesting to discover whether dispersion in risk across surviving managers would suffice to explain this result.

funds for the period 1974–1988, while the latter study considers a much larger sample of 728 mutual funds for the period 1976–1988, 258 of which survived for the entire period. The major conclusions of the two studies are similar. Performance persists.[3]

While the experimental designs and data of these studies differ considerably, the generic results may be illustrated on Tables 1 and 2. The relationship between successive three-year growth equity fund risk-adjusted total returns for the period 1976–1987 is documented in Table 1. The 2 × 2 contingency tables show the frequency with which managers who performed in the top half of all managers [on a Jensen (1968) α risk-adjusted basis] for a given three-year interval also performed in the top half in the subsequent three-year interval. For every period studied, the results are similar. If a manager wins in the first three years, the probability is greater than 50 percent that the manager will win in the second three years. These results are also statistically significant in at least two of the three successive three-year intervals.[4]

Goetzmann and Ibbotson (1991) report contingency tables similar to those given in Table 1 for a variety of time periods and performance horizon intervals. The data on which Table 1 is based are similar to those of the Hendricks, Patel, and Zeckhauser (1991) study. An alternative approach is to regress second-period Jensen's α's against first-period Jensen's α's. A significantly positive slope coefficient is evidence of persistence. The result of this exercise is presented in Table 2. The results correspond with those reported in Table 1. The evidence of persistence is strongest in the first and third subperiod of the data. Hendricks, Patel, and Zeckhauser (1991) suggest computing the returns on a self-financing portfolio strategy, a methodology they

[3] Note that the Kritzman (1983) and Dunn and Theisen (1983) results apply to pension fund money managers, while the other studies that indicate persistence all refer to mutual funds. Representatives from Frank Russell Company and other pension fund consulting companies indicate that efforts to replicate the mutual fund persistence results using pension fund data have to this date been unsuccessful. Part of the reason for this difference might be that mutual fund returns are measured after fees, while pension fund returns typically are measured before commissions (see note 6).

[4] One has to be a little careful interpreting the statistical significance of the χ^2 values. The identification of managers as winners or losers is actually ex post. For this reason, we expect to find the winners-following-winners result at least 50 percent of the time. This ex post conditioning also implies that the standard χ^2 tests (with or without the Yates 2 × 2 continuity correction) will be misspecified. Fortunately, an alternative statistic, the cross-product ratio (given as the ratio of the product of the principal diagonal cell counts to the product of the off-diagonal counts in the 2 × 2 table), has well-known statistical properties. Statisticians prefer the cross-product ratio (or measures closely related to it) because it simultaneously provides a test of the hypothesis that the two classifications are independent, as well as giving a measure of the dependence [Bishop et al. (1975, p. 373ff.)] In the present case, row and column, sums of each 2 × 2 contingency table are fixed because of ex post conditioning. Thus, the winner–winner cell count determines all other cell counts, and is distributed as the hypergeometric distribution conditional on row and column counts. Thus, the p-value of the cross-product ratio statistic is given by the sum of hypergeometric probabilities of cell counts at least as great as the observed winner–winner count [Agresti (1990, p. 60)]. This is known as Fisher's *exact test.*

Table 1
Two-way table of growth managers classified by risk-adjusted returns over successive intervals 1976–1987

Winners and losers defined relative to performance of median manager

	1979–1981 winners	1979–1981 losers	
1976–1978 winners	44	19	63
1976–1978 losers	19	44	63
	63	63	126
			χ^2 = 19.84 (p = .0)
			χ^2 (Yates correction) = 20.40 (p = .0)
			Cross-product ratio = 5.36 (p = .0)

	1982–1984 winners	1982–1984 losers	
1979–1981 winners	35	33	68
1979–1981 losers	33	35	68
	68	68	136
			χ^2 = 0.12 (p = .732)
			χ^2 (Yates correction) = 0.12 (p = .732)
			Cross-product ratio = 1.12 (p = .432)

	1985–1987 winners	1985–1987 losers	
1982–1984 winners	52	25	77
1982–1984 losers	25	52	76
	77	76	153
			χ^2 = 18.35 (p = .0)
			χ^2 (Yates correction) = 18.74 (p = .0)
			Cross-product ratio = 4.24 (p = .0)

This table is derived from total returns on growth equity mutual funds made available by Ibbotson Associates and Morningstar, Inc. Risk-adjustment is the Jensen (1968) α measure relative to total returns on the S&P 500 Index. Each cell represents the number of funds in the sample that share the characteristic defined by the row and the column. For example, the number of funds that were in the top half of mutual funds over the 1976–1978 period and were subsequently also in the top half of mutual funds over the 1979–1981 period may be found in the first row and first column of the upper 2 × 2 table. The χ^2 and χ^2 (Yates correction) refer to standard χ^2 test statistics for independence, where Yates refers to Yates 2 × 2 continuity correction. The cross-product ratio is the ratio of the product of principal diagonal cell counts to the product of the off-diagonal counts. Where (as in this case) the row and column sums are determined ex post the p-value can be inferred from the hypergeometric distribution of the upper left-hand cell count in the 2 × 2 table (Fisher's *exact test*).

attribute to Grinblatt and Titman (1989). The portfolio weights are proportional to the deviation of prior performance measures from the mean performance measure across managers. The performance measure of such a portfolio is a measure of persistence. This measure is computed in Table 2. The results are qualitatively similar to ones reported by Hendricks, Patel, and Zeckhauser (1991).

These results of course require careful interpretation. It is tempting to conclude from the type of results reported in Tables 1 and 2 that "hot hands" exist among mutual fund managers. Actually, the methodologies are silent on whether the persistence relates to positive or

Table 2
Regression-based measures of persistence in performance, 1976–1987

Cross-section regression approach[1]

Period January 1976–December 1981:

$$\hat{\alpha}_2 = .0885 + .4134\hat{\alpha}_1$$
$$\quad\ (5.38)\quad (6.47)$$
$$R^2 = 2.53;\ n = 126$$

Period January 1979–December 1984:

$$\hat{\alpha}_2 = -.0831 + .0070\hat{\alpha}_1$$
$$\quad\ (-3.69)\quad (0.07)$$
$$R^2 = .000;\ n = 134$$

Period January 1982–December 1987:

$$\hat{\alpha}_2 = -.0753 + .3052\hat{\alpha}_1$$
$$\quad\ (-6.53)\quad (5.28)$$
$$R^2 = .156;\ n = 153$$

Time-series self-financing portfolio approach[2]

Period January 1977–December 1987:

$$r_{pt} = .0018 - .0078r_{mt}$$
$$\quad\ (2.88)\quad (-.61)$$
$$R^2 = .003;\ n = 132$$

(t-values in parentheses)

[1] Jensen's α is computed for the sample of funds described in Table 1 for each of four three-year subperiods of data starting in 1976–1978. Each panel reports results from the cross-section regression of performance measures on prior performance measures. The first panel gives results from the regression of Jensen's α measures estimated on the basis of data for the period January 1979–December 1981 on similar measures estimated for the period January 1976–December 1978.

[2] This corresponds to the measure employed by Hendricks, Patel, and Zeckhauser (1991) with four quarter evaluation and holding periods. For each year starting in 1976, Jensen's α measures are computed. The deviation of these measures from their mean corresponds to a self-financing portfolio, which is then applied to excess returns on funds measured for the subsequent year. The portfolio is updated at the end of each year. The regression reports results from the time-series regression of the resulting monthly excess returns on market excess returns. The intercept corresponds to a performance measure for this portfolio strategy.

negative performance. This is most readily apparent in Table 1 when we observe that the row and column sums are specified ex post given the sample of money managers. In other words, given the row and column sums and the "winner–winner" cell count, the "loser–loser" count is simply the residual. Given the "loser–loser" count, the "winner–winner" is the residual. When we measure risk-adjusted performance relative to zero (Table 3), we find that persistence can just as easily relate to negative performance as it does to positive performance. Sometimes (1976–1981) good performance is rewarded by subsequent good performance. "Hot hands" are evident. Sometimes (1982–1987) it is the case that bad performance is punished by further bad news. This result is also apparent examining the intercepts of the cross-section regressions reported in Table 2. Results reported in Table 4 indicate that the persistence of poor performance serves to explain some but not all of the results reported in the previous tables. This table gives regression-based measures of persistence excluding those managers who experienced negative average Jensen's α for the

Table 3
Two-way table of growth managers classified by risk-adjusted returns over successive intervals 1976–1987

Winners and losers defined relative to zero risk-adjusted performance measure

	1979–1981 winners	1979–1981 losers	
1976–1978 winners	88	11	99
1976–1978 losers	16	11	27
	104	22	126

$\chi^2 = 12.92$ ($p = .0$)
χ^2 (Yates correction) $= 11.14$ ($p = .001$)
Cross-product ratio = 5.50

	1982–1984 winners	1982–1984 losers	
1979–1981 winners	42	72	114
1979–1981 losers	4	18	22
	46	90	136

$\chi^2 = 2.87$ ($p = .09$)
χ^2 (Yates correction) $= 3.13$ ($p = .08$)
Cross-product ratio = 2.62

	1985–1987 winners	1985–1987 losers	
1982–1984 winners	20	34	54
1982–1984 losers	15	84	99
	35	118	153

$\chi^2 = 9.49$ ($p = .002$)
χ^2 (Yates correction) $= 9.15$ ($p = .002$)
Cross-product ratio = 3.29

This table is derived using the same data as that reported in Table 2. Risk-adjustment is the Jensen's α measured relative to total returns on the S&P Index. Winners and losers are defined relative to Jensen's α measure of zero. For example, the number of funds that experienced a positive α over the 1976–1978 period and subsequently experienced a positive α over the 1979–1981 period may be found in the first row and first column of the upper 2 × 2 table. The χ^2 and χ^2 (Yates correction) refer to standard χ^2 test statistics for independence, where Yates refers to Yates 2 × 2 continuity correction. The cross-product ratio is the ratio of the product of principal diagonal cell counts to the product of the off-diagonal counts.

entire period 1976–1987. The results are similar to those reported in Table 3. The significance of apparent persistence has fallen. However, both the cross-section and the self-financing portfolio results indicate that there is still statistically significant evidence that performance persists for at least part of the period.

The persistence of negative performance is not surprising. Negative performance can persist where a subset of managers are immune from periodic performance review and where it is difficult to short sell shares of mutual funds.[5] It can be only institutional reasons such as

[5] In fact, Hendricks, Patel, and Zeckhauser provide little reliable evidence of "hot hands." Using either the value-weighted or the equal-weighted CRSP index benchmark, there is no significant

Table 4
Regression-based measures of persistence in performance, 1976–1987 (excluding poor performers)

Cross-section regression approach[1]

Period January 1976–December 1981:

$$\hat{\alpha}_2 = \underset{(5.48)}{.1463} + \underset{(3.13)}{.2736}\hat{\alpha}_1$$
$$R^2 = .113;\ n = 79$$

Period January 1979–December 1984:

$$\hat{\alpha}_2 = \underset{(1.04)}{.0317} - \underset{(-1.55)}{.1815}\hat{\alpha}_1$$
$$R^2 = .029;\ n = 82$$

Period January 1982–December 1987:

$$\hat{\alpha}_2 = \underset{(-3.09)}{-.0334} + \underset{(.81)}{.0521}\hat{\alpha}_1$$
$$R^2 = .008;\ n = 88$$

Time-series self-financing portfolio approach[2]

Period January 1977–December 1987:

$$r_{pt} = \underset{(2.15)}{.0008} - \underset{(-.20)}{.0015}r_{mt}$$
$$R^2 = .000;\ n = 132$$

(t-values in parentheses)

This table is intended to show the effect that different standards of performance review might have on measures of persistence in returns. The procedures and data are the same as those presented in Table 2, with the exception that managers are excluded whose average value of Jensen's α is negative over the entire period for which data is available.

[1] Jensen's α is computed for the sample of funds described in Table 1 for each of four three-year subperiods, of data starting in 1976–1978, excluding those funds that performed poorly over the entire period. Each panel reports results from the cross-section regression of performance measures on prior performance measures.

[2] This corresponds to the measure employed by Hendricks, Patel, and Zeckhauser (1991) with four quarter evaluation and holding periods, excluding poor performing funds.

these that allow a fund with sustained poor performance to survive.[6] It is the persistence of positive returns that would be remarkable, if true. The problems of interpretation caused by the ex post definition of winners and losers suggests that the results may also be sensitive to the most obvious source of ex post conditioning: survival.

It is clear that all managers depicted in the 2 × 2 tables have passed the market test, at least for the successive three-year periods. We have no data for the managers who did not survive. If the probability of

persistence of positive performance. The only benchmark for which they find any statistically significant evidence of persistence in positive performance is a self-created benchmark consisting of an equal-weighted average of returns on the mutual funds in their sample.

[6] Hendricks, Patel, and Zeckhauser (1991) give the example of the 44 Wall Street funds that survived the period 1975–1988 with a negative annual α of -1.90 (relative to the value-weighted CRSP index) and -4.27 (relative to the equal-weighted CRSP index). One potential explanation for the persistence of negative performance might be that mutual fund data compute returns after fees but before sales and load charges. The negative performance may simply reflect the persistence of high fees.

survival depends on past performance to date, we might expect that the set of managers who survive will have a higher ex post return than those who did not survive. Managers who take on significant risk and lose may also have a low probability of survival. This observation suggests that past performance numbers are biased by survivorship; we only see the track record of those managers who have survived. This does not suggest, however, that performance persists. If anything, it suggests the reverse. If survival depends on cumulative performance, a manager who does well in one period does not have to do so well in the next period in order to survive. Certainly, this survivorship argument cannot explain results suggested by Table 1. Moreover, there is a general perception that the survivorship bias effect cannot be very substantial. In a recent study, Grinblatt and Titman (1989) report that the survivorship effect accounts for only about 0.1 to 0.4 percent return per year measured on a risk-adjusted basis before transaction costs and fees. We shall see that the survivorship bias in mean excess returns is small in magnitude relative to a more subtle, yet surprisingly powerful, survival bias that implies persistence in performance.

A manager who takes on a great deal of risk will have a high probability of failure. However, *if* he or she survives, the probability is that this manager took a large bet and won. High returns persist. If they do not persist, we would not see this high-risk manager in our sample.[7] Note that this is a total risk effect; risk-adjustment using β or other measure of nonidiosyncratic risk may not fully correct for it. To illustrate this effect, observe in Table 3 that the additional 10 firms that come into the database in 1979–1981 are all ex post successful. The average value of residual risk (0.0323) for the new entrants is significantly greater than that of the population of managers (0.0242), with a t-value of 2.02. The new entrants who survived took on more risk and were successful.

The magnitude of the persistence will depend on the precise way in which survivorship depends on past performance and whether there is any strategic risk management response on the part of surviving money managers.[8] The intent is to show that the apparent

[7] Hendricks, Patel, and Zeckhauser (1991) argue that because fund data is eliminated from their database as the fund ceases to exist or is merged into other funds, their sample is free of survivorship bias effects. However, all funds considered at each evaluation point survived at least until the end of an evaluation period that could extend from one quarter to two years. They are excluded from the analysis *subsequent to* the evaluation period. The numerical example given in Section 2 of this article matches this experimental design, and provides a counterexample to a presumption of freedom from survivorship bias effects. The results of such a study would be free of survival bias only if it can be established that the probability of termination or elimination from the sample is unrelated to performance. However, Hendricks, Patel, and Zeckhauser indicate (note 5) that, in fact, funds that go under do quite poorly in the quarter of demise.

[8] We show in the Appendix that the effect is mitigated somewhat where cumulative performance

persistence of performance documented in Tables 1 and 2 is not necessarily any indication of skill among surviving managers.

To the extent that survivorship depends on past returns, ranking managers who survive by realized returns may induce an apparent persistence in performance. Survivorship implies that managers will be selected according to total risk. One way of explaining the Table 1 results is to observe that the set of managers studied represent a heterogeneous mix of management styles. Each management style is characterized by a certain vector of risk attributes. By examining the survivors, we are really only looking at those styles that were ex post successful. It may appear that one resolution of this problem is to concentrate on only one defined management style. There are two problems with this approach. In the first instance, we have to be careful to define the style sufficiently broadly that there are more than a few managers represented. In the second instance, we may exacerbate the effect if our definition of manager style is synonymous with taking high total risk positions.

We only observe the performance of managers who survive performance evaluations. The purpose of this article is to examine the extent to which this fact is sufficient to explain the magnitude of persistence we seem to see in the data. In Section 1, we examine the relationship between total risk differentials and survivorship-induced persistence in performance. In Section 2, we present some numerical results that show that a very small survivorship effect is sufficient to generate a strong and significant appearance of dependence in serial returns. We conclude in Section 3.

1. Relationship between Volatility and Returns Induced by Survivorship

There are many possible quite complex sample selection rules. We will look at the implications of one class of these rules. Our purpose in this section is to demonstrate that sample survivorship bias is a force that can lead to persistence in performance rankings. For simplicity, assume all distributions are atomless. Our tool is the following lemma.

---

rather than one-period performance is used as a survival criterion. The analysis of a strategic response is beyond the scope of this article. A possible strategic response is for surviving managers who are subject to the same survival criterion to converge in residual risk characteristics. The results in the next section require only that the ranking of managers by residual risk be constant. This kind of strategic response would also tend to mitigate the effect. This analysis is complicated by the fact that survival criteria are not necessarily the same for all managers.

Lemma. *Assume*

(i) x, y independent random variables,
(ii) $\Pr(x \geq 0) = \Pr(y \geq 0) > 0$,
(iii) $\Pr(x > a) \geq \Pr(y > a)$, $\quad \forall\ a \geq 0$, with strict inequality for some a.

It follows that

$$\Pr[x > y \mid x,y > 0] > \tfrac{1}{2}.$$

Before proving the lemma, note that its conditions are satisfied by x and y if they are both normal with mean zero, and if x has a higher variance than y. More generally, for both x and y with mean zero, it is sufficient if x is distributed as λy, where $\lambda > 1$.

Proof. Let F_x and F_y be the respective cumulative distributions of x and y, and let G_x and G_y be the reverse cumulants

$$G_x \equiv 1 - F_x, \qquad G_y \equiv 1 - F_y.$$

Now,

$$\begin{aligned}
&\Pr(x > y \mid x,y > 0) \\
&\quad = [\Pr(x,y > 0)]^{-1} \Pr(x > y,\ x > 0,\ y > 0) \\
&\quad = [G_x(0)\, G_y(0)]^{-1} \int_0^\infty \int_u^\infty dF_x(z)\ dF_y(u) \\
&\quad = [G_x(0)\, G_y(0)]^{-1} \int_0^\infty G_x(u)\ dF_y(u) \\
&\quad \geq [G_x(0)\, G_y(0)]^{-1} \int_0^\infty G_y(u)\ dF_y(u) \\
&\quad = [G_x(0)\, G_y(0)]^{-1} \{-\tfrac{1}{2} G_y^2(u) \,|_0^\infty\} \\
&\quad = \frac{1}{2} \left\{ \frac{G_y(0)}{G_x(0)} \right\} = \frac{1}{2},
\end{aligned}$$

with strict inequality if $G_x > G_y$ on some set of positive measure. ■

The following corollary establishes that the result generalizes to cases where x and y represent nonzero mean random variates.

Corollary 1. *Let ϵ_x and ϵ_y satisfy the conditions of the lemma and let*

$$x = f + \epsilon_x \qquad \text{and} \qquad y = f + \epsilon_y;$$

then

$$\Pr[x > y \mid \epsilon_x,\epsilon_y > 0] > \tfrac{1}{2}.$$

Proof. Immediate. ■

For the following set of results, let x and y be two random variables drawn from a family of distributions indexed by a spread parameter $\sigma \in [0, \infty)$. The family has the property that

$$G(0 \mid \sigma) = G(0)$$

and $(\forall\ a \geq 0)\ \sigma' > \sigma$ implies that

$$G[a \mid \sigma'] \geq G[a \mid \sigma].$$

Corollary 2. *It follows that*

$$\Pr[\sigma_x > \sigma_y \mid x > y;\ x,y > 0] > \tfrac{1}{2}.$$

Proof. From the lemma

$$\Pr[x > y \mid \sigma_x > \sigma_y;\ x,y > 0] > \tfrac{1}{2},$$

and, therefore, by initial symmetry and Bayes' theorem,

$$\begin{aligned}
&\Pr[\sigma_x > \sigma_y \mid x > y;\ x,y > 0] \\
&\quad = \frac{\Pr[x > y \mid \sigma_x > \sigma_y;\ x,y > 0]\Pr[\sigma_x > \sigma_y \mid x,y > 0]}{\Pr[x > y \mid x,y > 0]} \\
&\quad > \frac{1}{2}\left(\frac{1/2}{1/2}\right) = \frac{1}{2}.
\end{aligned}$$

■

The next two corollaries verify that if one random variable exceeds another in one observation period, it is likely to do so in other periods.

Corollary 3. *Let x and y be independent and (unconditionally) identically distributed with unknown spreads drawn from the family described above. It follows that*

$$\Pr[x_2 > y_2 \mid x_1 > y_1;\ x_1, y_1, x_2, y_2 > 0] > \tfrac{1}{2}.$$

Proof. For notational ease we will omit the ubiquitous conditioning on $x_i,y_i > 0$. From Bayes' theorem,

$$\begin{aligned}
&\Pr[x_2 > y_2 \mid x_1 > y_1] \\
&\quad = \Pr[x_2 > y_2 \mid \sigma_x > \sigma_y,\ x_1 > y_1]\ \Pr[\sigma_x > \sigma_y \mid x_1 > y_1] \\
&\qquad + \Pr[x_2 > y_2 \mid \sigma_x \leq \sigma_y,\ x_1 > y_1]\ \Pr[\sigma_x \leq \sigma_y \mid x_1 > y_1]
\end{aligned}$$

$$= \Pr[x_2 > y_2 \mid \sigma_x > \sigma_y] \Pr[\sigma_x > \sigma_y \mid x_1 > y_1]$$
$$+ \Pr[x_2 > y_2 \mid \sigma_x \leq \sigma_y] \Pr[\sigma_x \leq \sigma_y \mid x_1 > y_1],$$

since σ is sufficient statistic.

By the a priori symmetry of x and y,

$$1 = \Pr[x_2 > y_2 \mid \sigma_x \leq \sigma_y] + \Pr[x_2 \leq y_2 \mid \sigma_x \leq \sigma_y]$$
$$= \Pr[x_2 > y_2 \mid \sigma_x \leq \sigma_y] + \Pr[x_2 \geq y_2 \mid \sigma_x \geq \sigma_y].$$

From the lemma we know that

$$\Pr[x_2 > y_2 \mid \sigma_x > \sigma_y] \equiv \tfrac{1}{2} + p, \qquad p > 0,$$

and from Corollary 2 we know that

$$\Pr[\sigma_x > \sigma_y \mid x_1 > y_1] \equiv \tfrac{1}{2} + q, \qquad q > 0.$$

Hence,

$$\Pr[x_2 > y_2 \mid x_1 > y_1]$$
$$= (\tfrac{1}{2} + p)(\tfrac{1}{2} + q) + (1 - (\tfrac{1}{2} + p))(1 - (\tfrac{1}{2} + q))$$
$$= \tfrac{1}{2} + 2pq > \tfrac{1}{2}. \qquad \blacksquare$$

Corollary 4. *The conditions are the same as for Corollary 3 except that the unknown spread parameter is not constant, although ranking by volatility is preserved, that is,*

$$\sigma_{x_1} > \sigma_{y_1} \Rightarrow \sigma_{x_2} > \sigma_{y_2}.$$

Proof. Immediate, given

$$\Pr[x_2 > y_2 \mid \sigma_{x_2} > \sigma_{y_2}] = \Pr[x_2 > y_2 \mid \sigma_{x_1} > \sigma_{y_1}] \equiv \tfrac{1}{2} + p, \quad p > 0. \qquad \blacksquare$$

This concludes our analytic verification of the relation between volatility and returns in a sample that is truncated by survivorship bias. To say whether this effect is larger or smaller than the natural tendency for regression to the mean depends on the exact sample selection rules.

If the selection rule in a two-period model is $x_1 + x_2 > 0$ (and/or $x_1 > 0$), then we have verified the tendency for one fund to persist in outperforming the other if its volatility is higher. However, with the rule $x_1 + x_2 > 0$, there is another opposing force. In particular, if $x_1 > y_1 > 0$, then x does not have to pass so extreme a hurdle in the next period, and we are likely to have $x_2 < y_2$ if there is no dispersion in volatility across managers (see the Appendix). We avoided this problem in Corollary 3 by only conditioning on $x_1, y_1 >$

0 and not on $x_1 + x_2 > 0$ and $y_1 + y_2 > 0$. This two-period selection effect would tend to counterbalance the variance-induced apparent persistence of returns. The net effect will depend on assumptions made about the distribution of returns absent the selection effects.

Another mitigating factor would arise if we allow managers to adjust their portfolio policies to adjust risk levels on survival. While the above results require only that the ranking by risk be constant from one period to the next, the extent of the persistence will depend on the way in which survival affects the differences in risk across managers.

It is clear that the magnitude of the persistence in returns will depend on assumptions that are made about the precise nature of survivorship, the distribution of returns across managers, and the way in which portfolio policies of managers evolve through time. In the next section, we shall examine a simple numerical example that demonstrates that very mild survivorship criteria are sufficient to induce strong persistence in performance for a reasonable specification of the distribution of returns across managers.

2. A Numerical Illustration of the Magnitude of Induced Persistence in Returns

To examine the numerical magnitude of the persistence in performance induced by survivorship, annual returns were generated for a cross section of managers. The moments of the distribution of returns are chosen to match those observed in the data, although it is assumed that manager returns are serially uncorrelated. There is a natural presumption that persistence in observed returns implies that manager returns are predictable. The purpose of this experiment is to provide a reasonable counterexample to this presumption. While there may be many factors that are in fact responsible for the persistence in performance, a simple survivorship argument suffices to explain the magnitude of persistence we observe in the data.

Conditional on systematic risk measure β_i and nonsystematic risk σ_i, annual returns are generated from

$$R_{it} = r_f + \beta_i(R_{mt} - r_f) + \epsilon_{it},$$

where the annual Treasury bill rate r_f is taken to be 0.07 and the annual equity risk premium is assumed to be normal with mean 0.086 and standard deviation (SD) 0.208 corresponding to the Ibbotson and Sinquefeld (1990) numbers for the period 1926–1989. The idiosyncratic term ϵ_{it} is assumed to be distributed as normal with mean zero and SD σ_i.

The managers are defined by their risk measures β_i and σ_i. It is difficult to know what are reasonable values for these parameters. If observed mutual fund data is truncated in possibly complex ways by survivorship, then that data may yield biased measures of the parameters. If, on the other hand, it is held that this truncation is of a second order of importance, then, given this maintained hypothesis, the cross-sectional distribution of the parameters will give some measure of the underlying dispersion of risk. For the purpose of this experiment, it is assumed that β_i is distributed in the cross section of managers as normal with mean 0.95, and SD 0.25 corresponding to the cross-sectional distribution of β observed in the Goetzmann and Ibbotson sample of money managers.[9]

The distribution of nonsystematic risk across managers is functionally dependent on β. Closet index funds with β's close to unity typically have very low values of nonsystematic risk, whereas managers whose β's deviate from the market tend to be less well diversified. This suggests a relationship between nonsystematic risk and β approximated by[10]

$$\sigma_i^2 = k(1 - \beta_i)^2.$$

[9] As seen from Table 5, a 5 percent performance cut will lead to an increase in the average β of about 5 percent. The increase is due solely to the truncation in the cross-sectional distribution of β. This is an important caveat in interpreting Table 5 to imply a calibration of survival measures. There is a more subtle issue here. If there is a performance cut, ordinary least squares will not be appropriate. Beta should be estimated taking account of the fact that the distribution of residuals is truncated by survival. Assuming that the truncation by survival occurs on a quarterly basis, and that the minimum observed return among survivors (relative to the market) defines the point at which the residuals are truncated, it is possible to estimate a truncated regression model for the data described in Table 1. The measure of β was not sensitive to truncation; however, the measure of residual risk rose, on average, 2.5 percent. To the extent that our results depend on the distribution of the residual risk across managers, this represents another caveat to the results reported in Table 5.

[10] This proportional relationship does not only capture the apparent segmentation of mutual funds into closet index funds characterized by a β of unity and a low residual risk, and less well-diversified funds with β's less than or greater than unity. It also matches the empirical regularity that suggests that residual risk is an increasing function of the absolute difference of portfolio β from unity [e.g., Black, Jensen, and Scholes (1972)]. This relationship also follows for size-ranked portfolios and managed funds [Connor and Korajczyk (1991), Elton et al. (1993, Table 6)]. The constant of proportionality, k, was chosen so that the cross-sectional average R^2 matches the average value of .90 for the Goetzmann and Ibbotson sample. This value also corresponds to the available data. For the 438 money managers for whom Goetzmann and Ibbotson have data for the period 1984–1988, a regression of residual risk on the deviation of β from unity yields the following:

$$\begin{aligned}\hat{\sigma}_i^2 = \underset{(4.928)}{.000374} + \underset{(-3.827)}{.00012}\,(1 - \hat{\beta}_i) + \underset{(13.014)}{.005294}\,(1 - \hat{\beta}_i)^2,\\ R^2 = .360, \qquad N = 438\end{aligned}$$

(t-values in parentheses). To account for the possibility that this relationship may be an artifact of leptokurtosis in fund manager returns, β and residual risk are estimated on the basis of alternate-month returns [for a discussion of the related issue of skewness-induced correlation of sample mean returns and volatility, see Roll and Ross (1980)]. Assuming that the cross-sectional distribution of returns is truncated by the lowest observed return in that month (the survivor), a truncated regression approach applied to the same data yields a coefficient on the squared term of .003552 (t-value 15.325) with intercept and linear terms statistically insignificant. These values expressed on an annualized basis correspond closely to the value for k, 0.05349, used in the simulation experiments.

The value of k chosen in the simulation experiment is 0.05349, which is the value that ensures that the average R^2 across managers is 0.90, given the distribution of β and the assumed variance of the equity risk premium.

The experiment proceeds as follows. For each of 600 managers, a value of β_i is chosen. This defines a measure of nonsystematic risk σ_i given the assumed relationship between the two parameters. Four annual returns are drawn for each manager according to the assumed return generating process. For each of four years, the worst performing managers are eliminated from the group.[11] Four-year returns are computed for each of the managers that survive this sequential cut, and contingency tables corresponding to those of Table 1 and regression-based measures of persistence corresponding to Tables 2 and 4 are calculated.

The results of this experiment will obviously depend on the severity of the cut. In a base case analysis, no managers are cut. In a second scenario, only the bottom 5 percent of managers are cut in each year. In the third and fourth scenarios, the bottom 10 percent and 20 percent of managers are eliminated each year. The entire experiment is then repeated 20,000 times. In this way, we examine not only the expected frequency of persistence as a function of the selection criterion, but also the sampling properties of this persistence.

In our first exercise, we generate results corresponding to Table 1. Risk-adjusted performance measures are evaluated for each manager using Jensen's α, and the cumulated risk-adjusted returns are computed for the first two years and second two years. "Winners" and "losers" are defined relative to the risk-adjusted performance of the median manager in each two-year period. This experimental design follows closely the approach adopted by Hendricks, Patel, and Zeckhauser (1991) and Goetzmann and Ibbotson (1991), with the important exception that in constructing risk-adjusted returns, β is assumed known. Thus, we do not consider the possible complications that arise from the necessity to estimate this quantity.

The average values of the frequency of persistence in risk-adjusted return across the replications of this experiment, for different assumed cutoff points, are given in Table 5. When there is no truncation by survivorship, there is no apparent persistence of performance. However, when managers are excluded from the sample for performance reasons, there is evidence of apparent persistence in performance. The probability is greater than 50 percent that a manager who wins

[11] Consistent with Corollary 3, managers are truncated in the final year. Failure to truncate in the final year leads to a small decrease in the apparent persistence of performance in Table 5. However, the qualitative conclusions are not affected by this change.

Table 5
Two-way table of managers classified by risk-adjusted returns over successive intervals, a summary of 20,000 simulations assuming 0, 5, 10, and 20 percent cutoffs

	Second-period winners	Second-period losers
	No cutoff (n = 600)	
First-period winners	150.09	149.51
First-period losers	149.51	150.09
		Average χ^2 = 1.04
		Average cross-product ratio = 1.014
		Average cross-section t-value = −.004
		Average annual excess return = 0.0%
		Average β = 0.950
	5% cutoff (n = 494)	
First-period winners	127.49	119.51
First-period losers	119.51	127.49
		Average χ^2 = 1.64
		Average cross-product ratio = 1.159
		Average cross-section t-value = 2.046
		Average annual excess return = 0.44%
		Average β = 0.977
	10% cutoff (n = 398)	
First-period winners	106.58	92.42
First-period losers	92.42	106.58
		Average χ^2 = 3.28
		Average cross-product ratio = 1.366
		Average cross-section t-value = 3.356
		Average annual excess return = 0.61%
		Average β = 0.994
	20% cutoff (n = 249)	
First-period winners	71.69	53.31
First-period losers	53.31	70.69
		Average χ^2 = 7.13
		Average cross-product ratio = 1.919
		Average cross-section t-value = 4.679
		Average annual excess return = 0.80%
		Average β = 1.018

For each simulation, manager annual returns are drawn from the market model described in the text, allowing for a dispersion in β and nonsystematic risk in the cross section of managers. In each of the four years, managers who experience returns in the lowest percentile indicated by the cutoff value are excluded from the sample, and this experiment is repeated 20,000 times. Thus, the numbers in the first 2 × 2 table give the average frequency with which the 600 managers fall into the respective classifications. The second panel shows the average frequencies for the 494 managers who survive the performance cut, while the third and fourth panels give corresponding results for 398 and 249 managers. For each simulation, the winners are defined as those managers whose average two-year Jensen's α measure was greater than or equal to that of the median manager in that sample. The average χ^2 refers to the average value of the standard χ^2 test statistic for independence (without Yates' correction) across the simulations, while the average cross-product ratio refers to the average value of the ratio of the product of principal diagonal cell counts to the product of the off-diagonal counts.

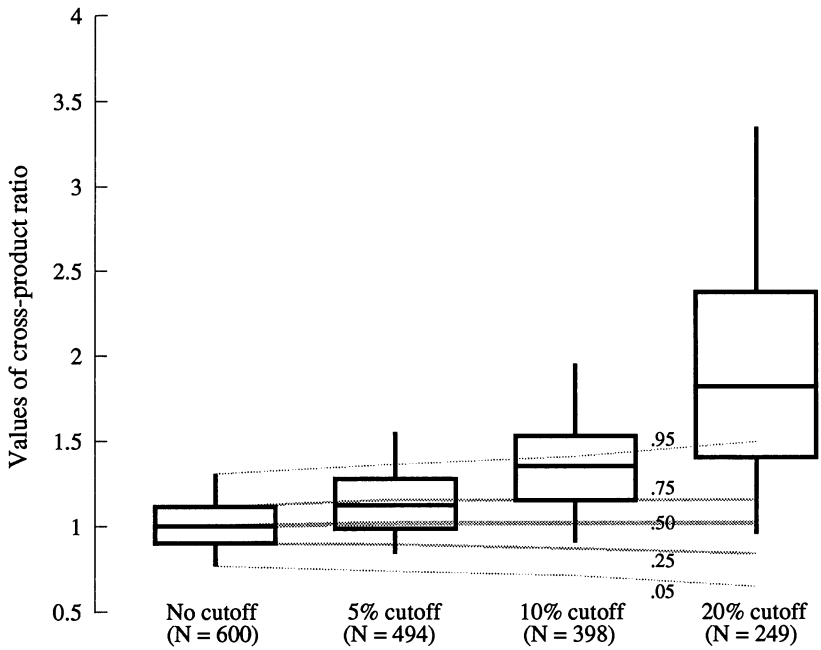

Figure 1
Boxplots of 20,000 simulated values of the cross-product ratio for different performance cutoff levels
The solid line within each box represents the median of the empirical distribution of cross-product ratios, whereas the box itself gives the interquartile range. The whiskers above and below each box give the 95th and 5th percentiles, respectively, of the empirical distribution. The gray lines give the stated fractiles of the theoretical distribution implied by the hypergeometric distribution of cell counts assuming independence.

in the first period will also win in the second. This probability increases with the extent to which the sample is truncated by survivorship. The effect of truncation is also evident in the increase in the average value of χ^2 and cross-product ratio statistics with the degree to which managers are excluded from the sample for performance reasons. The effect is particularly marked when we regress performance measures in the second two-year period on similar measures computed for the first two-year period. With no cutoff, the mean t-value for the slope coefficient of this regression is zero. However, both the mean and the median t-values are in excess of 2 with just a 5 percent performance cut. This means that on the basis of a cross-section regression of successive α's, we would reject the hypothesis of no persistence at least 50 percent of the time!

The effect of truncation on the distribution of test statistics for dependence is quite marked. In Figure 1, we display the boxplots of simulated values of the cross-product ratio for different performance cutoff levels. This solid line within each box represents the median of the empirical distribution of cross-product ratios, whereas the box itself gives the interquartile range. The whiskers above and below each box give the 95 and 5 percentiles, respectively, of the empirical distribution. For comparison purposes, we provide the theoretical distribution implied by the hypergeometric distribution of cell counts implied by independence. When there is no truncation by survivorship, the distribution of this statistic is well specified. However, when managers have to survive a performance cut, there appears to be evidence of short-term dependence in performance. When only 5 percent of managers are cut in each of the first three years, the cross-product ratio is too high relative to its theoretical distribution assuming independence. With a 10 percent performance cutoff each year, three quarters of the time the test statistic lies above the median of its theoretical distribution. Even a small degree of truncation by survivorship will induce an unacceptably high probability of false inference of persistence in performance.

It might be argued that the apparent persistence we observe in these simulation experiments is some artifact of the way in which the test statistics have been computed. After all, the cross-product ratio is not widely used in the finance literature. Hendricks, Patel, and Zeckhauser (1991) argue in favor of a t-value statistic based on the returns computed on the basis of a self-financing portfolio strategy where the portfolio weights are proportional to the deviation of performance measures from the average performance measure across managers. The results of this kind of approach are given in Table 2, and the simulation results are presented in Figure 2. Note that this test statistic is, if anything, more misspecified under a performance cut than is the cross-product ratio. Given that this is true even in the special case where we know precisely the β of the self-financing portfolio, and can compute the theoretical variance of the performance measure, we would expect the performance of the statistic under realistic experimental conditions to be much worse.

It is important to note that truncation by survivorship may imply an apparent persistence in performance without significantly affecting average risk-adjusted returns. As we observed before, Grinblatt and Titman (1989) find that survivorship bias can account for only about 0.1 to 0.4 percent return per year. Table 5 shows the average risk-adjusted returns for managers who survive the various performance cuts. While there are substantial differences in average risk-adjusted return between managers who did well and poorly in the successive two-year periods, the net effect of survivorship bias on average risk-adjusted returns for all managers in the sample is very small and corresponds to about 0.4 to 0.6 percent per year on a risk-adjusted basis for the 5 to 10 percent cutoff examples. The corresponding

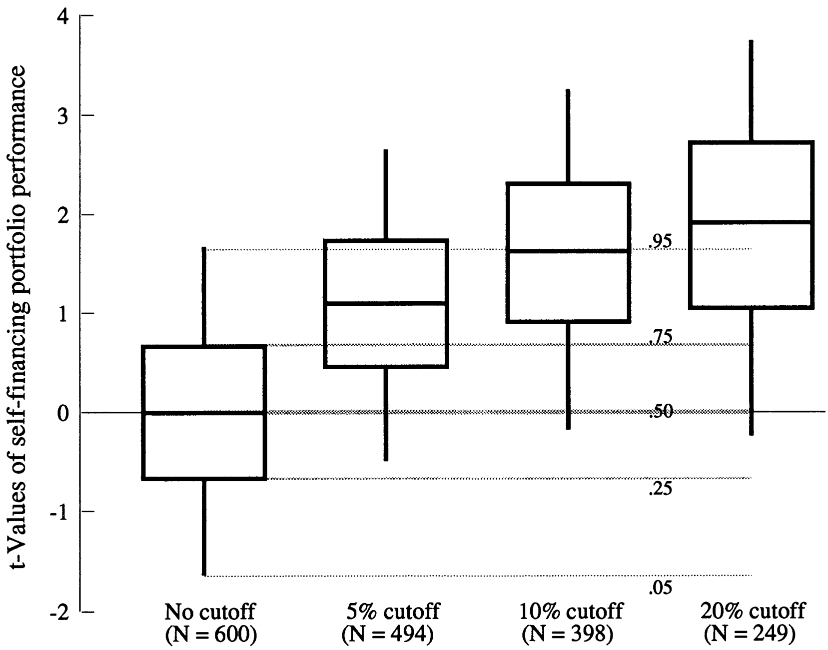

Figure 2
Boxplots of t-values associated with 20,000 simulated values of the self-financing portfolio performance measure
The solid line within each box represents the median of the empirical distribution of t-values associated with the Jensen's α of the self-financing portfolio strategy described in the text (assuming β known), whereas the box itself gives the interquartile range. The whiskers above and below each box give the 95th and 5th percentiles, respectively, of the empirical distribution. The gray lines give the stated fractiles of the theoretical distribution implied by the null hypothesis of a zero performance measure.

number is 0.8 percent for the 20 percent cutoff. These numbers do not differ significantly from those reported by Grinblatt and Titman. It would appear from the results reported in Table 5 that truncation of raw returns is compensated for by a corresponding truncation in the cross-sectional distribution of β, leading to no net effect on average risk-adjusted returns.

Of course, it might be said that these results are something of a straw man. After all, the example assumes that manager performance is evaluated on a total return basis. Actually, the apparent persistence in performance is even stronger than that reported in Table 5 if managers are terminated for low α, representing risk-adjusted returns. This result is implied by Corollary 1 above. In this example, what is important is not the dispersion across managers of total risk, but rather the dispersion of residual risk. This suggests that it may be possible to mitigate some of the survival effect by simply standardizing per-

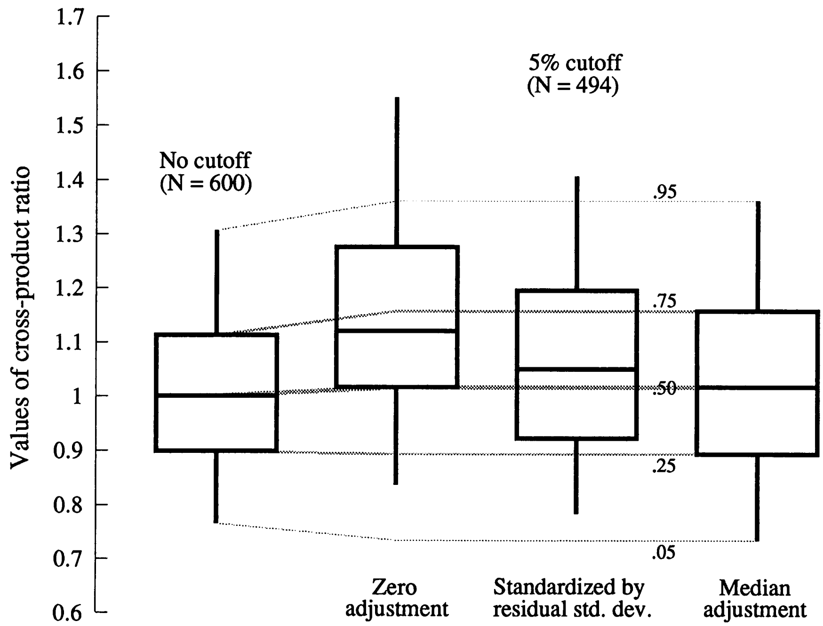

Figure 3
Boxplots of 20,000 simulated values of the cross-product ratio showing the effect of different adjustments for survivorship bias with a 5 percent performance cutoff
Zero adjustment refers to the boxplot given in Figure 1, where there is a 5 percent performance cutoff. Standardized by the residual standard deviation refers to the cross-product ratio calculated on the basis of defining "winners" and "losers" relative to the median appraisal ratio. Median adjustment corrects the median for the fact that the distribution of appraisal ratios is truncated by survivorship. This correction is described in the text.

formance measures by the residual standard deviation.[12] In fact, classifying managers into winner and loser categories by α measured in units of residual risk does reduce the apparent persistence in Table 5. This reduces the dispersion of measures of persistence but does not eliminate the survivor-induced bias. To eliminate the bias we need to adjust excess returns to account for the fact that the median excess return will be greater than zero by virtue of survivorship.[13] The

[12] This application of the *appraisal ratio,* originally due to Treynor and Black (1972), was suggested to us by William Sharpe. A recent study to examine the properties of this ratio is Lehmann and Modest (1987).

[13] If the performance cut occurs at the 10th percentile of the unconditional distribution of manager returns, the median of the truncated distribution will occur at the 55th percentile of the unconditional distribution. To correct for survivorship bias, we first compute the fractile p of the distribution of excess returns for the particular manager that corresponds to the minimum observed return. The quantity q given as the $(1 - (1 - p)/2)$th quantile of the distribution of excess returns is the median excess return induced by survivorship. The median adjustment given in Figure 2 is obtained by subtracting q computed for each manager from that manager's annual excess returns. Obviously, this adjustment is highly sensitive to the assumptions made about the distribution of excess returns for each manager, and about the effect of past performance on survival.

effect of these separate adjustments on the apparent survivorship effect is illustrated in Figure 3.

The numerical example is unrealistic in at least one important respect. In common with the results reported in Table 1, it assumes that the excess of returns of managers are cross-sectionally uncorrelated. Of course, there are patterns of performance related to styles of management, and we would expect excess returns to be correlated. In fact, in the sample period covered by Table 1, the intra-manager correlation of excess returns can reach as high as .98. This will exacerbate the effect if the pattern of intercorrelation depends on measures of risk. One high-risk manager surviving will increase the chance that other high-risk managers will also survive.

The degree of intercorrelation among managers does indeed appear to be functionally dependent upon β and residual risk.[14] Results of an experiment where the cross-sectional correlation of excess returns corresponds to the Goetzmann and Ibbotson study[15] are presented in Figure 4. The cross-correlation effect is sufficiently strong to cause a false inference of persistence even in the absence of a performance cut. Where there are performance cuts, this effect is considerably exacerbated.

Figure 4 indicates that the cross-product ratio test statistic is seriously misspecified. To obtain some idea of the order of magnitude, recall that a cross-product ratio of 4 corresponds to a contingency table where the cell counts on the diagonal are twice the off-diagonal terms. With a 5 percent performance cut, apparent dependence of this magnitude will be observed at least five percent of the time. It is important to note that the simple cross-section regression approach is also misspecified. The upper 95 percentile of the resulting test statistic, 1.65, is exceeded 32.9 percent of the time with no performance cut. With a 5 percent performance cut, this percentile is

[14] Two hundred fifty money managers in the Goetzmann and Ibbotson database were ranked according to β. The average β and intracorrelation of performance measures was computed for each of 20 groupings by β. As a purely descriptive measure, the average intracorrelations were related to β as follows:

$$\hat{\rho}_{ij} = \underset{(2.04)}{.558} - \underset{(-2.66)}{.732}(\hat{\beta}_i + \hat{\beta}_j) + \underset{(4.29)}{1.216}(\hat{\beta}_i \times \hat{\beta}_j),$$
$$R^2 = .3580, \quad N = 190$$

(t-values in parentheses). If the true correlation matrix corresponds to this regression equation, it is a simple exercise in matrix algebra to show that the distribution of residual returns is a two-factor structure, with factor loadings and idiosyncratic variances given as analytic functions of the β and β-product terms. This two-factor structure is used to generate Figures 3 and 4 in the text. As an aside, the same exercise in linear algebra shows that principal components will be an ineffective control for cross-sectional dependence, since the idiosyncratic variances of residual returns will be a quadratic function of β. Using principal components assumes the idiosyncratic variances are constant in the cross section.

As discussed earlier, this result is subject to the important caveat that the residual covariance matrix and β are estimated without regard to the possible effects of survival on the cross-sectional distribution of these parameters and on the distribution of residuals.

[15] This uses the two-factor structure described in note 14.

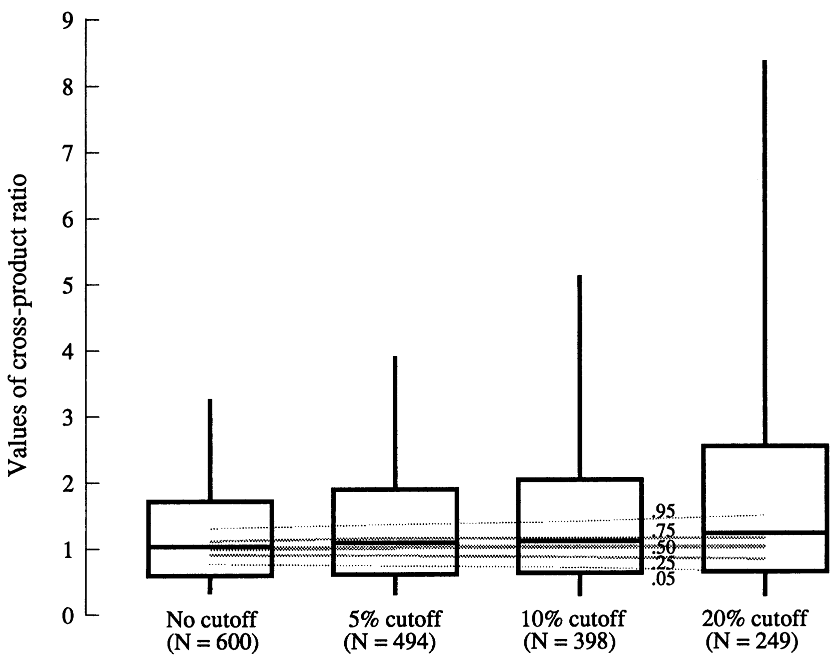

Figure 4
Boxplots of 20,000 simulated values of the cross-product ratio showing the effect of cross-correlation in performance measures
This figure corresponds to Figure 1, where the cross-correlation of performance measures matches that of the Goetzmann and Ibbotson (1991) sample. The procedure used to induce this level of cross-correlation is described in the text.

exceeded 54.39 percent of the time. The median value of the distribution of t-value statistics is 2.09.

The theoretical distribution assumes the performance measures are uncorrelated in the cross section of managers. Where we induce cross-sectional correlation into the performance measures, with no performance cut the cross-product ratio is unbiased but the variance is far greater than the theoretical distribution would imply. The cross-section regression approach, which imposes far more restrictive assumptions on the process generating sequential returns, is even more seriously affected. One concludes that the combination of dependence in the cross-section distribution of returns with truncation by performance might be sufficient to explain the results reported in Table 1.

Where there is cross-sectional dependence, the median adjustment is not well specified, although it does represent an improvement over the unadjusted statistics, as indicated in Figure 5. This adjustment assumes the excess returns are independent in the cross section. It is sensitive to violations of this assumption. While it is possible to

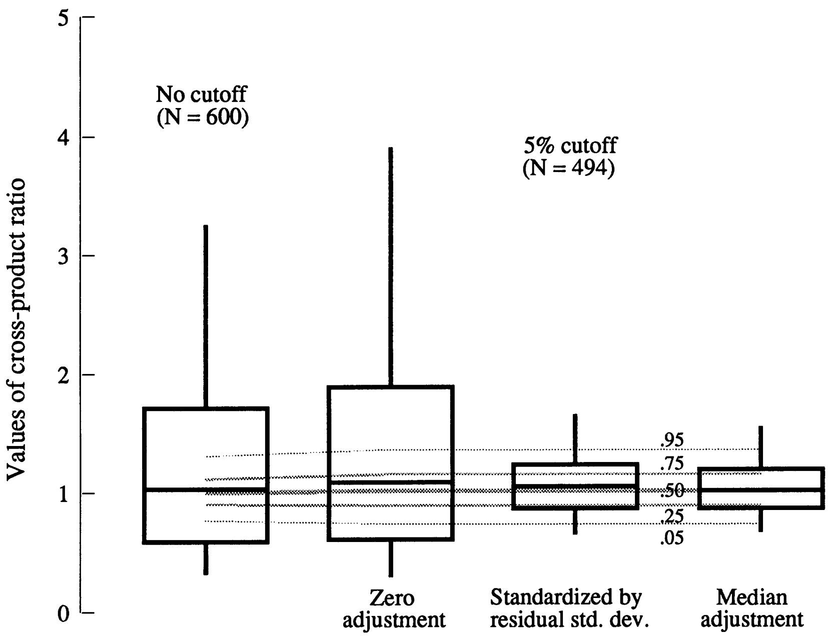

Figure 5
Boxplots of 20,000 simulated values of the cross-product ratio showing the effect of cross-correlation on alternative adjustments for survivorship bias
This figure corresponds to Figure 3, where the cross-correlation of performance measures matches that of the Goetzmann and Ibbotson (1991) sample. The procedure used to induce this level of cross-correlation is described in the text.

conceive of an exact adjustment based on the order statistics assuming dependence in manager excess returns, it is interesting to note that the simple residual standard deviation adjustment does at least as well as the median adjustment. This simple measure requires no information about the magnitude of the performance cut. The result suggests the conjecture that the simple prescription of normalizing performance numbers by residual standard deviations may represent a reasonably robust performance statistic.[16]

[16] To illustrate the likely effects of normalizing performance measures by residual standard deviation, results reported in Table 4 were recomputed using this approach. The α and standard deviation measures are estimated using a truncated regression approach, where each month's return is assumed truncated from below by the return of the lowest manager in the group (the survivor). All measures of persistence are now statistically insignificant. The cross-section t-value for 1976–1981 falls from 3.13 to 1.77. The t-value for the self-financing portfolio approach performance now measures 1.76, whereas before it was 2.16. Two important caveats are in order. The result is sensitive to assumptions made about the way in which past performance influences survival. One could use information on firms that leave the sample to derive an explicit model for survival to construct a more powerful test. This would appear to be a standard application of the censored regression methodology were it not for the model-specific heteroskedasticity implied [see, e.g., Hurd (1979)]. Among other things, such a model would also need to account for cross-sectional correlation of manager performance. The second caveat is that these tests assume manager returns are independent through time. We would not expect such tests to be powerful against an alternative that allows manager returns to be autocorrelated absent the survival effect.

3. Conclusion

We show that truncation by survivorship gives rise to an apparent persistence in performance where there is dispersion of risk among money managers. Standard risk-adjustment technology, which adjusts for single-factor β risk, may not suffice to correct for this effect. A numerical example shows that this effect can give rise to a substantial probability that statistical tests based on risk-adjusted return data will give rise to the false inference that there is in fact dependence in security returns.

Our findings in this article are suggestive of implications beyond performance measurement. Where inclusion in a sample depends in part on rate of return, survivorship bias will lead to obvious biases in first and second moments and cross moments of return, including β. What is not so obvious is that this effect will induce a spurious relationship between volatility and return. This has implications for empirical tests of asset pricing models and in particular for studies of so-called anomalies.[17] It also has implications for studies of post-event performance of firms that survive significant corporate events. Current work examines whether survival bias of the kind reported here may suffice to explain the puzzling post-earnings-drift phenomenon first noted by Ball and Brown (1968) if there is dispersion of residual risk among those firms that survive into the post-earnings sample.[18]

Whether these results suffice to explain the strength of results reported by Goetzmann and Ibbotson (among others) is at this point an open question. We have shown that truncation by survival has a measurable impact on the observed returns of those managers who survive the performance cut. Clearly, the magnitude of the effect will depend on the fraction of managers who in fact survive the performance cut.[19] Furthermore, the numerical example was based on the dispersion of risk measures for managers *who survived.* In addition

[17] Since small firms are less diverse in their activities, we do not find it surprising that the residual risk for such firms is greater than for larger firms. The results of this article would suggest a survival-induced correlation between size and average (risk-adjusted) return.

[18] For more discussion on the post-earnings drift phenomenon, see Foster, Olson, and Shevlin (1984) and Bernard and Thomas (1989, 1990). For a discussion of survival bias effects as they relate to measures of accounting earnings, see Salamon and Smith (1977) and Ball and Watts (1979).

[19] Inspecting various annual issues of the Wiesenberger Investment Companies Service *Investment Companies* periodical, we find that for the period 1977–1987 the apparent attrition rate given as the fraction of equity fund managers who simply disappear from coverage, merge, or change their names ranges from 2.6 percent in 1985 to 8.5 percent in 1977, an average attrition rate of 4.8 percent. This average attrition is very close to the 5 percent attrition found by Grinblatt and Titman (1989). However, this number is very much a lower bound on the true attrition rate. To the extent that the number of equity funds increases through time, we should expect that the attrition rate will also increase.

it is assumed that survival depends on four annual reviews based solely on returns measured over the previous year.

To calibrate the magnitude of the possible bias, we need to know how the characteristics of managers who survive differ from other managers, and the role of past performance in determining which managers survive. Clearly, cumulative performance must have a role in this process. The strength of the apparent persistence evident in Table 1 seems to broadly correlate with periods of high volatility in the markets; market conditions may also play a role. As Hendricks, Patel, and Zeckhauser (1991) indicate, in the period 1974–1988, a subset of poorly performing managers appears to be immune from performance review. This factor alone seems to explain most of the apparent persistence in their study. These represent important issues for future research. Until they are resolved, it is difficult to devise a simple adjustment to standard performance measures that will correct for this survivorship bias.

Finally, the simulation results lead to the conjecture that the simple prescription of normalizing performance measures by the residual standard deviation might provide a performance measure that is relatively robust to this source of misspecification. However, there is an important caveat. These experiments assume that the true parameters of the process are known to the investigator. The task of estimating the risk measures in the presence of a potential performance cut and of designing a performance measure that corrects for the resulting apparent persistence in performance is the subject of ongoing research.

Appendix

In the text, we demonstrate that with the selection rule conditioning on early performance, there is a tendency for performance to persist. In this section, we show that if the selection rule conditions on overall (two-period) performance, then there is a tendency for performance reversal. The net effect of these two forces must be resolved empirically.

The basic problem we want to consider is

$$\Pr[x_2 > y_2 \mid x_1 > y_1, c],$$

where

$$c = \{x_1 + x_2 > 0, y_1 + y_2 > 0\}.$$

From Bayes' theorem

$$\Pr[x_2 > y_2 \mid x_1 > y_1, c] = \frac{\Pr[x_2 > y_2, x_1 > y_1, c]}{\Pr[x_1 > y_1, c]},$$

and again, by Bayes' theorem,

$$\Pr[x_1 > y_1, c] = \Pr[x_1 > y_1, c] = \Pr[x_1 > y_1 \mid c]\Pr[c].$$

For the purposes of this section we will ignore the possibility of dispersion in the spread parameter and assume x and y have independent and identical distributions. It follows that

$$\Pr[x_1 > y_1 \mid c] = \tfrac{1}{2}.$$

If we further assume that the distributions are symmetric about the origin, then

$$\Pr[c] = \tfrac{1}{4}, \qquad \Pr[x_1 > y_1, c] = (\tfrac{1}{2})(\tfrac{1}{4}) = \tfrac{1}{8}.$$

We now have the following result.

Lemma. *Under the above conditions*

$$\Pr[x_2 > y_2 \mid x_1 > y_1, x_1 + x_2 > 0, y_1 + y_2 > 0] < \tfrac{1}{2}.$$

Proof.

$$\begin{aligned}
&\Pr[x_2 > y_2, x_1 > y_1, c] \\
&\quad \equiv \Pr[x_2 > y_2, x_1 > y_1, x_1 + x_2 > 0, y_1 + y_2 > 0] \\
&\quad = \Pr[x_2 > y_2, x_1 > y_1, y_1 + y_2 > 0] \\
&\quad = \int_{-\infty}^{\infty} \int_{y_1}^{\infty} dF \left\{ \int_{-y_1}^{\infty} \int_{y_2}^{\infty} dF\, dF_{y_2} \right\} dF_{y_1} \\
&\quad = \int_{-\infty}^{\infty} G(y_1) \int_{-y_1}^{\infty} G(y_2)\, dF_{y_2}\, dF_{y_1} \\
&\quad = \int_{-\infty}^{\infty} G(y_1) \left\{ -\frac{1}{2} G^2(y_2) \Big|_{-y_1}^{\infty} \right\} dF_{y_1} \\
&\quad = \frac{1}{2} \int_{-\infty}^{\infty} G(y_1) G^2(-y_1)\, dF_{y_1} \\
&\quad = \frac{1}{2} \int_{-\infty}^{\infty} G(y_1)[1 - G(y_1)]^2\, dF_{y_1} \quad \text{(by symmetry)} \\
&\quad = \tfrac{1}{2}(-\tfrac{1}{2}G^2 + \tfrac{2}{3}G^3 - \tfrac{1}{4}G^4) \big|_{-\infty}^{\infty} \\
&\quad = \tfrac{1}{2}\left\{\tfrac{1}{2} - \tfrac{2}{3} + \tfrac{1}{4}\right\} = \tfrac{1}{24}.
\end{aligned}$$

Hence,

$$\Pr[x_2 > y_2 \mid x_1 > y_1, c] = \frac{1/24}{1/8} = \frac{1}{3} < \frac{1}{2}. \qquad \blacksquare$$

This is the tendency for reversal in the absence of any inferences about volatility from returns. It is clear, by continuity, that if we permitted a small disparity in ex post spreads for x and y, this effect would still dominate. However, as the possibility of spreads is increased, the persistence described in the text also increases. In theory and in practice, which effect is dominant depends on both the exact form of the selection rules and the potential dispersion of the spread parameter.

References

Agresti, A., 1990, *Analysis of Ordinal Categorical Data,* Wiley, New York.

Ball, R., and P. Brown, 1968, "An Empirical Examination of Accounting Income Numbers," *Journal of Accounting Research,* 6, 159–178.

Ball, R., and R. Watts, 1979, "Some Additional Evidence on Survival Biases," *Journal of Finance,* 34, 1802–1808.

Bernard, V., and J. Thomas, 1989, "Post–earnings-announcement Drift: Delayed Price Response or Risk Premium," *Journal of Accounting Research,* 27, 1–36.

Bernard, V., and J. Thomas, 1990, "Evidence that Stock Prices Do Not Fully Reflect the Implications of Current Earnings for Future Earnings," *Journal of Accounting and Economics,* 13, 305–340.

Bishop, Y., S. Fienberg, and P. Holland, 1975, *Discrete Multivariate Analysis: Theory and Practice,* MIT Press, Cambridge, Mass.

Black, F., M. Jensen, and M. Scholes, 1972, in M. Jensen (ed.), *Studies in the Theory of Capital Markets,* Praeger, New York.

Connor, G., and R. Korajczyk, 1991, "The Attributes, Behavior and Performance of U.S. Mutual Funds," *Review of Quantitative Finance and Accounting,* 1, 5–26.

Cowles, A., 1933, "Can Stock Market Forecasters Forecast?" *Econometrica,* 1, 309–325.

Dunn, P., and R. Theisen, 1983, "How Consistently Do Active Managers Win?" *Journal of Portfolio Management,* 9, 47–50.

Elton, E., and M. Gruber, 1989, *Modern Portfolio Theory and Investment Analysis,* Wiley, New York.

Elton, E., M. Gruber, and J. Rentzler, 1990, "The Performance of Publicly Offered Commodity Funds," *Financial Analysts Journal,* 46, 23–30.

Elton, E., M. Gruber, S. Das, and M. Hlavka, 1993, "Efficiency with Costly Information: A Reinterpretation of Evidence for Managed Portfolios," forthcoming in *Review of Financial Studies.*

Foster, G., C. Olsen, and T. Shevlin, 1984, "Earnings Releases, Anomalies, and the Behavior of Security Returns," *Accounting Review,* 59, 574–603.

Goetzmann, W., and R. Ibbotson, 1991, "Do Winners Repeat? Patterns in Mutual Fund Behavior," working paper, Yale School of Organization and Management.

Grinblatt, M., and S. Titman, 1988, "The Evaluation of Mutual Fund Performance: An Analysis of Monthly Returns," Working Paper 13-86, John E. Anderson Graduate School of Management, University of California at Los Angeles.

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

Hendricks, D., J. Patel, and R. Zeckhauser, 1991, "Hot Hands in Mutual Funds: The Persistence of Performance, 1974–88," working paper, John F. Kennedy School of Government, Harvard University.

Hurd, M., 1979, "Estimation in Truncated Samples When There Is Heteroscedasticity," *Journal of Econometrics,* 11, 247–258.

Ibbotson, R., and R. Sinquefeld, 1990, *Stocks Bonds Bills and Inflation: 1990 Yearbook,* Ibbotson Associates, Chicago.

Ippolito, R., 1989, "Efficiency with Costly Information: A Study of Mutual Fund Performance 1965–1984," *Quarterly Journal of Economics,* 104, 1–23.

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

Kritzman, M., 1983, "Can Bond Managers Perform Consistently?" *Journal of Portfolio Management,* 9, 54–56.

Lehmann, B., and D. Modest, 1987, "Mutual Fund Performance Evaluation: A Comparison of Benchmarks and Benchmark Comparisons," *Journal of Finance,* 21, 233–265.

Roll, R., and S. Ross, 1980, "An Empirical Investigation of the Arbitrage Pricing Theory," *Journal of Finance,* 35, 1073–1103.

Salamon, G., and E. D. Smith, 1977, "Additional Evidence on the Time Series Properties of Reported Earnings Per Share," *Journal of Finance,* 32, 1795–1801.

Treynor, J., and F. Black, 1972, "Portfolio Selection Using Special Information under the Assumptions of the Diagonal Model with Mean Variance Portfolio Objectives and without Constraints," in G. P. Szego and K. Shell (eds.), *Mathematical Models in Investment and Finance,* North-Holland, Amsterdam.