

Inference about Survivors

Robert F. Stambaugh *

Abstract

This study explores inference about assets that have survived by avoiding poor performance. The greater is the commonality across assets in prior uncertainty about parameters, the more an asset's inferred expected return should depend on its having survived. If there is no commonality, a surviving asset's average return can possess substantial sampling bias while nevertheless equaling the appropriate conditional expected return. Survival bias as typically computed generally makes too severe an adjustment to survivors, unless one assumes that expected returns on all assets, dead or alive, are equal to one common value that is completely unknown before observing returns data.

JEL Classifications: C11, C34, G12

Keywords: survival bias, survival effects

* The Wharton School, University of Pennsylvania, and NBER. I am grateful for comments from Klaas Baks, Kobi Boudoukh, Stephen Brown, Richard Evans, Neal Galpin, Chris Jones, Shmuel Kandel, Andrew Metrick, Ľuboš Pástor, Guofu Zhou, workshop participants at Carnegie Mellon University, HEC, Imperial College, INSEAD, London Business School, New York University, Stanford University, Texas A&M University, the University of California at Berkeley, the University of Michigan, and the University of Texas, as well as participants in the 2003 NBER Asset Pricing Meetings and the 2003 SBFSIF Conference.

Author contact info: Finance Department, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6367, 215-898-5734, stambaugh@wharton.upenn.edu.

1. Introduction

Hedge funds and other investment vehicles often cease to exist as separate entities due to poor performance. The fact that an asset is currently available for investment often means that its past performance has been sufficient to keep it alive. Moreover, the return histories for non-surviving assets can be unobservable. Hedge funds, for example, can elect well into their lives to begin supplying their returns to performance-reporting services. The returns of funds with performance insufficient to induce voluntary reporting, or even to allow survival, are thus not available when making inferences about funds currently available for investment. Before investing in an asset, an investor might wish to estimate its expected return. Suppose the investor makes this inference by relying on historical data available for that asset and other survivors. How should this inference account for the fact that these assets have survived while others have not?

The survival process plays a key role in such inferences, and the nature of that process is rarely known with precision. Even with a known survival process, however, inferences about a survivor depend critically on a number of other issues. Foremost among these is the commonality across assets in one's prior uncertainty about the joint probability distribution of asset returns. Commonality in such uncertainty means that knowledge about the parameters relevant to one asset supply information about the parameters relevant to another. For example, commonality implies that learning (hypothetically) that one fund manager's true alpha is positive would produce an upward revision in beliefs about the true alphas of other managers. The stronger is this property of one's prior beliefs, the more important is conditioning on survival when making inferences about a survivor.

The term "survival bias" is often encountered in discussions of survival. In empirical studies, survival bias is typically computed as the average return difference between a set of surviving funds and a broader set that includes dead funds. A few of the many studies applying this general approach to mutual funds include Grinblatt and Titman (1989), Malkiel (1995), and Brown and Goetzmann (1995), and the biases generally range from less than 0.5% to nearly 2% per annum. Applications to hedge funds include Brown, Goetzmann, and Ibbotson (1999), Fung and Hsieh (2000), and Liang (2000), with computed biases of roughly 2% to 3%. Is there an appropriate use of such a survival bias value when estimating a particular surviving fund's expected return? One might be inclined to subtract this bias value from the fund's average return, but this value gives a sensible correction for survival only if one assumes that the true expected returns on all funds, living and dead, are identical. In that setting, one would also pool all available data across funds, giving no special role to

a fund's own average return when trying to infer that fund's expected return.

An assumption of identical expected returns (or identical alphas) across assets is often employed in analyses of survival issues, such as in Brown, Goetzmann, Ibbotson, and Ross (1992), Carhart, Carpenter, Lynch, and Musto (2002), and Cochrane (2005). Such an assumption is the extreme case of commonality in prior uncertainty about assets' expected returns. As shown here, conditioning on survival generally has the largest effect on inferences about a survivor when expected returns on all assets are assumed to equal a common value about which there is no prior information. More likely, it would seem, is a scenario in which one would have difficulty being sure that all assets have the same expected return while having no idea a priori about what that expected return might be. When the commonality in prior uncertainty falls short of the extreme case, a surviving asset's own sample average supplies at least some unique information about that asset's expected return, and the traditional survival bias is too large an adjustment.

Various forms of commonality in unexpected returns across assets also play an important role in determining the extent to which conditioning on survival affects inferences about surviving assets. Such commonality can arise from prior beliefs that surviving assets have similar sensitivities to a set of specified factors. It can also arise as positive correlations among factor-adjusted returns on different assets, or what might be termed style commonality. In the complete absence of commonality in prior parameter uncertainty as well as unexpected returns, survival essentially becomes irrelevant to making inferences about survivors. This special-case scenario corresponds to that analyzed by Baks, Metrick, and Wachter (2001), who make the seemingly counter-intuitive observation that conditioning on survival has no impact on inferences about surviving assets.

In statistical terms, "bias" typically refers to the difference between a statistic's expected value, in repeated samples, and the true value of the unknown parameter the statistic is intended to estimate. A worthwhile distinction can be drawn between a statistic's sampling bias and the role of that statistic in making conditional inference about the unknown parameter. Section 2 begins the analysis by drawing this distinction in a survival setting, borrowing the expositional "helicopter-tour" device used by Sims and Uhlig (1991) to make the same distinction in a unit-root analysis. An initial example demonstrates how, for an asset surviving a simple minimum-return threshold, the sample average return can be upward biased but nevertheless equal to the correct conditional inference about that asset's unknown true mean. The role of commonality across assets in one's parameter uncertainty is then demonstrated using the same framework, with one asset's survival depending on

outperforming another. The sample average still provides the correct conditional inference if one's uncertainty about each asset's mean is unrelated to the other but not when one believes the true means on the assets to be identical.

Section 3 presents a more formal Bayesian setting in which to analyze the problem of making inferences about the parameters of surviving assets conditional on data that do not include the returns on the non-surviving assets. Section 4 explains how a Markov-Chain Monte Carlo (MCMC) approach can be used to obtain posterior distributions for parameters of surviving assets, conditional on their having survived. In this study's general framework, the returns on non-survivors are treated as unobservable, motivated by the hedge-fund scenario noted above. An alternative setting with observable returns on non-survivors is also discussed. Section 5 presents a simple four-fund illustration in what corresponds loosely to a stylized hedge-fund setting, where some aspects of prior uncertainty are specified using an empirical Bayes procedure applied to a dataset of equity hedge funds. In order to simplify the analysis and focus on the aspects of prior uncertainty and commonality described at the outset, the number of non-survivors as well as the survival process are taken as known. The results illustrate the key role of commonality in prior uncertainty: the greater that commonality, the larger is the reduction in an asset's alpha when conditioning on its having survived. Also included is an example in which conditioning on survival actually pushes one's inference about an asset's alpha in the opposite direction. That paradoxical result occurs with commonality in the uncertainty about assets' factor sensitivities but little if any commonality in uncertainty about alphas. Section 6 considers the issue of style commonality, demonstrating its tendency to increase the importance of survival conditioning. Section 7 investigates the role of prior beliefs that relate alphas to residual risk, and section 8 concludes with a brief summary and some directions for future research.

2. A helicopter tour

Before proceeding to a more formal Bayesian framework, the study begins by examining inferences about a surviving asset in a simpler setting. Consider the problem of inferring the expected return on an asset whose cumulative realized return over a 120-month sample period exceeds a known survival criterion, to be specified below. The asset's return in each month is drawn independently from a normal distribution with a known standard deviation of 15% and an unknown mean of α . (All means and standard deviations of returns in these examples are expressed on an annualized basis.) The value of α is known to be drawn from

a uniform distribution on the $[-20\%, 20\%]$ interval.¹ Let a denote the asset’s sample mean return over a given sample. Relations between α and a can be viewed on a “helicopter tour” analogous to that provided by Sims and Uhlig (1991) for the coefficient of an autoregressive process.

As an initial case, suppose that the survival criterion is simply that the asset’s realized cumulative 120-month return exceeds a fixed threshold of -10%. The first step is to generate many samples, each containing 120 months, where each sample is generated using a new value of α drawn from the above uniform distribution. From those samples, retain only those in which the cumulative return realized exceeds the survival criterion. The joint frequency distribution of α and a from the retained samples is shown in the upper plot in Figure 1. The middle plot in Figure 1 slices the joint frequency distribution along a vertical plane parallel to the a axis at the value $\alpha = 5\%$, and that slice reveals the sampling distribution of a . The mean of that distribution is 6.3%, which implies an upward bias in a of 1.3% when $\alpha = 5$. As is well known, the sample mean a is subject to a survival bias, in that α is less than the expected value of a across repeated samples meeting the survival criterion.² The bottom plot in Figure 1 slices the joint distribution in the upper plot along a vertical plane parallel to the α axis at the value $a = 5\%$. This slice reveals the distribution of α conditional on observing a sample in which $a = 5\%$. The mean (and median) of this distribution is 5%, equal to the value of the sample mean a .³

The above example is perhaps the most transparent setting in which to contrast conditional inference and sampling bias in the presence of a survival criterion. An investor wishing to estimate the expected return in a manner that is correct when averaged across repeated samples would find a unsatisfactory for that purpose, given its positive sampling bias. At the same time, across all samples with a given sample mean a , the average α is equal to a . That is, an investor wishing to infer the value of α from a given sample at hand, the expected value of α conditional on that sample is simply a . In essence, if one conditions on the sample of the surviving asset’s returns, further noting that those returns meet the survival criterion provides no additional information. The sample is then summarized by a , the single sufficient statistic in this setting with normal returns and a known variance.

Next consider a relative survival criterion that depends on the performance of another

¹These finite truncation points, imposed for computational feasibility, should be viewed as approximations to $(-\infty, \infty)$. The effects of the truncation on the quantities reported below are negligible.

²Brown, Goetzmann, and Ross (1995) and Li and Xu (2002) analyze properties of sample paths conditioned on survival.

³Technically, the equality is only approximate, given the truncation of the uniform distribution from which α is drawn, but the approximation error is negligible here.

asset. In particular, the asset in question is selected from a set of two assets, where the asset with the greatest cumulative return over the 120-month sample period is the one that survives. Each asset's return is again generated from a normal distribution with an unknown mean and the same known standard deviation as before. In this setting there are two alternative assumptions about the mean returns on the two assets. Under the first assumption, the α 's on the two assets are different, drawn independently from the same uniform distribution as before. Under the second assumption, the α 's on the two assets are the same, given by a single draw from that uniform distribution. Figure 2 displays the joint distributions of the surviving asset's α and a under the alternative assumptions, where each plot is created by drawing many samples of returns on the two assets. Figure 3 displays plots in which each of the joint distributions in Figure 2 is sliced at the value $\alpha = 5\%$. As before, these slices reveal the sampling distribution of a at that value. In both cases, the mean of the sampling distribution for a exceeds α , by 0.9% when the two α 's are independent (upper plot) and by 2.7% when they are the same value (lower plot). The relative survival criterion again imparts an upward bias, in repeated samples, to the sample average return of the surviving asset.

Figure 4 displays the distributions obtained by slicing the joint distributions in Figure 2 at the value $a = 5\%$. In the upper plot, where the α 's on the two assets are drawn independently, the mean of α for the surviving asset is 5%. That is, the conditional mean of α is again given by the sample mean a . In essence, the cumulative return on the non-surviving asset serves like a randomly drawn threshold replacing the fixed threshold in the initial example. Since the α 's on the two assets are independent, that threshold is independent of the true expected return on the surviving asset. Therefore, knowing that the surviving asset's returns exceeded such a threshold provides no information about that asset's α beyond what is provided by the observed sample of returns.

A different result occurs in the second plot in Figure 4, where α 's on the surviving and non-surviving assets are drawn as a single value. The same relative survival criterion as before now produces a 2.7% difference between the sample average return and the conditional mean of α . Here, knowing the surviving asset outperformed another asset with the same α is useful information about that unknown value, beyond what is provided by the surviving asset's returns. In essence, knowing that the surviving asset outperformed another asset with the same expected return now leads one to infer it is more likely than not that the surviving asset's returns exceeded their expected value α . In the previous cases, no such inference is supported. Also recall from the lower plot in figure 3 that, in this case of complete commonality (i.e., equality) in the two assets' alphas, the sampling bias in a is also

2.7%. As noted earlier, the usual notion of survival “bias” corresponds most closely to an appropriate adjustment for conditional inference when alphas on different are known to be equal to a common unknown value.⁴

By now it is probably clear that the result $E(\alpha|a) = a$ does not hinge on having a fixed-return threshold as opposed to a relative-performance criterion. In the initial example with the survival criterion of beating a cumulative -10% return, no mention is made of another asset that failed to beat that threshold. Suppose there were such an asset. If the α on that other asset is drawn independently of the surviving asset’s α , then conditioning on the former asset’s not surviving has no effect on the conditional mean of the surviving asset’s α . On the other hand, if the non-surviving asset’s α is equal to that of the surviving asset, then conditioning on that asset’s not clearing the fixed-return threshold still implies $E(\alpha|a) < a$, as obtained under the same assumption with the relative-performance example above. Here again, one then knows that the surviving asset outperformed another asset with the same α , so it is more likely than not that the surviving asset’s returns exceeded that value.

The assumption that the α ’s on the surviving and non-surviving asset are the same is the limiting version of a more general setting in which the α ’s on the surviving and non-surviving assets are simply drawn with some degree of commonality. In general, such commonality implies that the expected value of a surviving asset’s α is less than that asset’s sample mean return a .

If the distribution from which the true α ’s are drawn is viewed as a Bayesian prior distribution, then the above conditional distribution of a surviving fund’s α given its a is simply the Bayesian posterior distribution of α . Consider, for example, the relative survival criterion used in Figures 2 through 4. A Bayesian decision maker whose prior for a given surviving asset’s α is uniform on the interval $[-20\%, 20\%]$ and independent of the α ’s of non-surviving assets would view that fund’s a as its predictive expected return, and the posterior distribution of the fund’s α corresponds to the first plot in Figure 4. The next section develops a more general Bayesian setting in which commonality across surviving and non-surviving assets in prior beliefs about unknown parameters gives rise to survival effects in posterior distributions.

⁴That 2.7% arises as the same value in both cases is not an exact implication but rather a close approximation, reflecting the fact that in this example the probability density of an asset’s return conditional on α and being the survivor is nearly scale invariant with respect to α . In essence, α and a each then serve as location parameters in the distribution of the other.

3. The model and prior beliefs

3.1. Stochastic setting and likelihood function

The model includes n risky assets in addition to a riskless asset. Without loss of generality, assume that these risky “assets” are zero-investment positions, so that their “returns” represent a spread between a risky and riskless return or a spread between two risky returns. Assume that k of the n assets serve as benchmarks in a factor-based pricing model. Let f_t contain the returns in period t on the k benchmark factors, and let r_t contain the payoffs on m ($= n - k$) non-benchmark assets. Define the multivariate regression,

$$r_t = \alpha + Bf_t + u_t, \quad (1)$$

where $E\{u_t|f_t\} = 0$,

$$\Sigma = \text{Cov}\{u_t, u_t'\}, \quad (2)$$

and u_t is distributed multivariate normal, independently and identically across t . Also define

$$\mu = E\{f_t\} \quad \text{and} \quad \Omega = \text{Cov}\{f_t, f_t'\}. \quad (3)$$

Let x_t denote the vector of all n asset returns, $x_t' = (r_t' \ f_t')$. The first and second moments of x_t are given by

$$E = E\{x_t\} = \begin{bmatrix} \alpha + B\mu \\ \mu \end{bmatrix} \quad (4)$$

and

$$V = \text{Cov}\{x_t, x_t'\} = \begin{bmatrix} B\Omega B' + \Sigma & B\Omega \\ \Omega B' & \Omega \end{bmatrix}. \quad (5)$$

The set of model parameters, denoted by θ , consists of $(\alpha, B, \Sigma, \mu, \Omega)$.

Let the $T \times m$ matrix R contain returns on the m non-benchmark assets over T periods, and let the $T_L \times k$ matrix F contain realizations of k factors. The history of the factors can exceed that of the non-benchmark returns, so that $T \leq T_L$.

3.2. Specifying priors

3.2.1. Commonality in alphas

The prior for a non-benchmark asset’s alpha, an element of α , has the form,

$$\alpha_i|\alpha_0 \sim N(\alpha_0, \sigma_\alpha^2), \quad \alpha_0 \sim N(\bar{\alpha}_0, \psi^2), \quad (6)$$

where $\bar{\alpha}_0$, σ_α , and ψ are specified constants, and the prior for α_i conditional on α_0 is independent across assets. A prior belief about the cross-sectional dispersion in assets' alphas (both survivors and non-survivors) is controlled by σ_α . Uncertainty about the "grand" mean of funds' alphas is controlled by ψ . Of course, for ψ large, the value specified for the prior grand mean, $\bar{\alpha}_0$, is unimportant. In essence, the larger is ψ relative to σ_α , the greater is the commonality across assets in uncertainty about alphas, and thus expected returns. The hierarchical prior in (6) has close parallels in parametric empirical Bayes approaches and the shrinkage estimation of James and Stein (1963). Often, in such settings, the value of σ_α is treated as an unknown hyperparameter and ψ is effectively set to infinity.⁵ In this study, both σ_α and ψ are set to a number of alternative values in order to assess the roles of both dimensions of prior uncertainty in making inferences about survivors.

Let α denote the $m \times 1$ vector containing the α_i 's. The unconditional prior for α implied by (6) is given by

$$\alpha \sim N(\iota_m \bar{\alpha}_0, \Phi_\alpha), \quad (7)$$

where

$$\Phi_\alpha = \sigma_\alpha^2 I_m + \psi^2 \iota_m \iota_m', \quad (8)$$

and I_m denotes the $m \times m$ identity matrix.

3.2.2. Commonality in factor sensitivities

Asset i 's sensitivities to the factors are contained in the vector b_i , the i -th column of B' . As with the prior for the α_i 's, the prior for the b_i 's includes a belief about the dispersion across assets in the sensitivity to a given factor as well as uncertainty about the grand mean of assets' sensitivities to that factor. For each asset i ,

$$b_i | b_0 \sim N(b_0, \Phi_{b|b_0}), \quad b_0 \sim N(\bar{b}_0, \Phi_{b_0}), \quad (9)$$

and, conditional on the grand mean b_0 , prior beliefs about the b_i 's are independent across assets. Let $b = \text{vec}(B')$. The unconditional prior density for b implied by (9) is given by

$$b \sim N(\iota_m \otimes \bar{b}_0, \Phi_b), \quad (10)$$

⁵See, for example, Efron and Morris (1973), Morris (1983), and Carlin and Louis (1996). For early applications to portfolio problems, see Jorion (1986) and Frost and Savarino (1986). After circulating the first draft of this paper in 2002, the author learned of independent work in progress by Chris Jones and Jay Shanken, later published as Jones and Shanken (2005), in which a prior of this form is employed. They specify σ_α^2 as uncertain as well, which they argue can be especially important as the number of assets becomes large. This study also differs from theirs by analyzing a setting in which the returns on the non-surviving assets are not observed.

where

$$\Phi_b = I_m \otimes \Phi_{b|b_0} + \iota_m \iota_m' \otimes \Phi_{b_0}. \quad (11)$$

3.2.3. Remaining specifications

In the initial examples, the prior for Σ is given by a non-informative inverted Wishart distribution. Specifically,

$$p(\Sigma) \propto |\Sigma|^{-\frac{\nu+m+1}{2}} \exp\left\{-\frac{1}{2} \text{tr} H \Sigma^{-1}\right\}, \quad (12)$$

where

$$H = (\nu - m - 1)s^2 I_m, \quad (13)$$

and the number of degrees of freedom, ν , is set to a small value ($m + 2$). Properties of the inverted Wishart distribution imply

$$E\{\Sigma\} = \frac{1}{\nu - m - 1} H = s^2 I_m. \quad (14)$$

The value of s^2 is specified using an empirical Bayes procedure (setting it equal to the average of the sample residual variances of the observed assets), although this choice is unimportant with the small value for ν .

The prior for the vector of factor means, μ , is multivariate normal,

$$p(\mu) \propto \exp\left\{-\frac{1}{2}(\mu - \bar{\mu})' \Phi_\mu^{-1}(\mu - \bar{\mu})\right\}. \quad (15)$$

and the prior for Ω is of the standard diffuse form,

$$p(\Omega) \propto |\Omega|^{-\frac{k+1}{2}}. \quad (16)$$

Define the $(k + 1) \times m$ matrix $A = (\alpha B)'$, and let $a = \text{vec}(A)$. Combining (7) through (10) gives

$$p(a) \propto \exp\left\{-\frac{1}{2}(a - \bar{a})' \Phi_a^{-1}(a - \bar{a})\right\}, \quad (17)$$

where

$$\bar{a} = \iota_m \otimes \begin{bmatrix} \bar{\alpha}_0 \\ \bar{b}_0 \end{bmatrix} \quad (18)$$

and

$$\Phi_a = I_m \otimes \begin{bmatrix} \sigma_\alpha^2 & 0 \\ 0 & \Phi_{b|b_0} \end{bmatrix} + \iota_m \iota_m' \otimes \begin{bmatrix} \psi^2 & 0 \\ 0 & \Phi_{b_0} \end{bmatrix}. \quad (19)$$

The joint prior for all parameters is given by

$$p(\theta) = p(a)p(\Sigma|a)p(\mu)p(\Omega) \quad (20)$$

where $p(a)$ is given in (17), $p(\Sigma)$ is given in (12), $p(\mu)$ is given in (15), and $p(\Omega)$ is given in (16).

The above setup for prior beliefs treats the m assets symmetrically in the sense that the priors for α_i and b_i in (6) and (10) are the same across assets. That is, the prior moments \bar{a} and Φ_a , take the forms in (18) and (19), which treat all assets identically. This structure is imposed for convenience and is easily relaxed. The methodology presented is unaffected if the prior moments \bar{a} and Φ_a are set to any arbitrary values (as long as Φ_a is a positive-definite matrix). One could instead assume, for example, that the structure in (6) pertains to only a subset of assets for a given $\bar{\alpha}_0$, σ_α , and ψ . The α_i 's for another subset of assets could then be assigned a different degree of commonality. One subset of assets, say a group of hedge funds, might have a larger value of ψ , reflecting uncertainty about the overall contribution of hedge fund managers, as a class. Another subset of assets might be a group of more traditional investment vehicles, where ψ for that group might be assigned a smaller value.

4. Conditioning on survival

4.1. General framework

Recall that the joint probability density for R and F , $p(R, F|\theta)$, depends on the vector of unknown parameters, θ . Consider a partitioning of the m non-benchmark (left-hand) assets into subsets of size m_1 and m_2 ($m = m_1 + m_2$), with the $T \times m$ matrix of returns partitioned as $R = [R_1 \ R_2]$. Note that the joint unconditional distribution for θ and R_2 can be written as

$$p(\theta, R_2) = p(R_2|\theta)p(\theta), \quad (21)$$

where the marginal distribution $p(\theta)$ corresponds to the prior in (20). Let s be an $m \times 1$ vector of 0's and 1's whose elements identify the subset of surviving assets. Now assume that the sample realization of s is such that the above set of m_2 assets are the non-survivors, so that the set of m_1 assets are the survivors, with observed histories.

Assumed that, conditional on realized returns and factors, survival does not depend on the unobserved parameters of the distribution for those quantities. That is,

$$p(s|R, F, \theta) = p(s|R, F). \quad (22)$$

This assumption about the survival process, also made by Baks, Metrick, Wachter (2001), seems reasonable. It essentially says that deaths of assets arise from decisions made on the basis of observed outcomes.

A Markov Chain Monte Carlo (MCMC) approach can be used to obtain the desired posterior distribution of the unknown parameters, $p(\theta|R_1, F, s)$. The likelihood function given the observed data (R_1 , F , and s) can be written as

$$\begin{aligned} p(R_1, F, s|\theta, R_2) &= p(s|R_1, R_2, F, \theta)p(R_1, F|\theta, R_2) \\ &= p(s|R_1, R_2, F)p(R_1, F|\theta, R_2) \\ &= p(s|R_1, R_2, F)\frac{p(R_1, R_2, F|\theta)}{p(R_2|\theta)}. \end{aligned} \quad (23)$$

Taking the product of (21) and (23) gives

$$\begin{aligned} p(\theta, R_2|R_1, F, s) &\propto p(\theta, R_2)p(R_1, F, s|\theta, R_2) \\ &= p(\theta)p(R_1, R_2, F|\theta)p(s|R_1, R_2, F) \end{aligned} \quad (24)$$

$$= p(\theta)p(R_1, F|\theta)p(R_2|R_1, F, \theta)p(s|R_1, R_2, F). \quad (25)$$

The full conditional posterior density for θ is, from (24),

$$p(\theta|R_1, F, s, R_2) \propto p(\theta)p(R_1, R_2, F|\theta), \quad (26)$$

which is the posterior for the problem in which both R_1 and R_2 are observed. The full conditional posterior density for R_2 is, from (25),

$$p(R_2|R_1, F, s, \theta) \propto p(R_2|R_1, F, \theta)p(s|R_1, R_2, F). \quad (27)$$

Drawing repeatedly from the full conditionals in (26) and (27), using an MCMC procedure, gives the desired marginal posterior distribution, $p(\theta|R_1, F, s)$.⁶ Details of the procedure are provided in the Appendix.

In the illustration analyzed in the following sections, survival is determined by a known threshold applied to a fund's cumulative return over an initial subperiod. This specification is adopted for simplicity, in order to focus on other aspects of the problem. In this case, survival is a non-random function of R_1 , R_2 , and F , so the density $p(s|R_1, R_2, F)$ in (27) is simply an indicator function of R_2 . A probabilistic survival rule could be specified instead, such as a hazard or probit model.⁷ In addition, the parameters in such models could either be treated as known or assigned a prior.

⁶A textbook treatment of MCMC approaches is provided by Chen, Shao, and Ibrahim (2000).

⁷Lunde, Timmermann, and Blake (1999) estimate a hazard model for mutual fund survival, and Brown and Goetzmann (1995) estimate a probit model.

4.2. Special case: no commonality

A special case of the above setting is where the likelihood function is assumed to obey

$$p(R_1, R_2, F|\theta) = p(R_1, F|\theta_{1,F})p(R_2|F, \theta_2), \quad (28)$$

and the prior density is of the form

$$p(\theta) = p(\theta_{1,F}, \theta_2) = p(\theta_{1,F})p(\theta_2). \quad (29)$$

In this case, the factors account for all of the dependence between R_1 and R_2 , and the prior for the parameters in the conditional distribution of R_2 is independent of the prior for the parameters in the joint distribution of R_1 and F . Substituting (28) and (29) into (24) gives

$$p(\theta_{1,F}, \theta_2|R_1, F, s) \propto p(\theta_{1,F})p(\theta_2)p(R_1, F|\theta_{1,F})p(R_2|F, \theta_2)p(s|R_1, R_2, F), \quad (30)$$

so

$$p(\theta_{1,F}|R_1, F, s) \propto p(\theta_{1,F})p(R_1, F|\theta_{1,F}), \quad (31)$$

and thus

$$p(\theta_{1,F}|R_1, F, s) = p(\theta_{1,F}|R_1, F). \quad (32)$$

That is, inferences about the parameters of the surviving assets are unaffected by conditioning on s . This special case corresponds to that in Baks, Metrick, and Wachter (2001). Their model satisfies the independence assumptions in (28) and (29). In that setting, they make the interesting observation that, with the assumption in (22) about the survival process, inferences about the surviving funds are not affected by conditioning on their having survived.

4.3. Observable non-survivor returns

In the general framework above, the returns on the non-surviving assets are assumed to be unobservable. Such a scenario typically occurs in the reporting of hedge fund returns. A fund often elects, well into its life, to begin reporting its returns to a performance-reporting service, at which point the returns are backfilled to the fund's inception. Fung and Hsieh (2000) refer to this phenomenon as "instant-history." As a result, the returns of a fund that fails to achieve initial performance favorable enough to induce voluntary reporting, or even to survive, are not available when making inferences about survivors. In other settings, however, it can be reasonable to assume that the data do not totally exclude assets that

fail to meet a survival criterion. Instead, the data can include returns on a comprehensive universe of all funds up to the point when they die.⁸ The mutual-fund database developed in Carhart (1997) and currently updated and supplied by CRSP is such an example.

When the non-survivor returns are observable, the earlier MCMC approach still applies but in a simpler fashion: as long as the survival process satisfies (22), the specific nature of that process is irrelevant. To see this, further partition R_2 into the subsets $R_{2,A}$ and $R_{2,B}$, where the first subset contains the returns observed before death and the second contains hypothetical after-death returns. Rewrite (21) as

$$p(\theta, R_{2,D}) = p(R_{2,D}|\theta)p(\theta), \quad (33)$$

and (23) as

$$\begin{aligned} p(R_1, R_{2,A}, F, s|\theta, R_{2,D}) &= p(s|R_1, R_{2,A}, F, R_{2,D}, \theta)p(R_1, R_{2,A}, F|\theta, R_{2,D}) \\ &= p(s|R_1, R_{2,A}, F)p(R_1, R_{2,A}, F|\theta, R_{2,D}) \\ &= p(s|R_1, R_{2,A}, F)\frac{p(R_1, R_{2,A}, R_{2,D}, F|\theta)}{p(R_{2,D}|\theta)}, \end{aligned} \quad (34)$$

again using (21) and the further simplification

$$p(s|R_1, R_2, F) = p(s|R_1, R_{2,A}, F), \quad (35)$$

since survival cannot depend on the hypothetical post-death returns. Taking the product of (33) and (34) gives

$$\begin{aligned} p(\theta, R_{2,D}|R_1, R_{2,A}, F, s) &\propto p(\theta, R_{2,D})p(R_1, R_{2,A}, F, s|\theta, R_{2,D}) \\ &= p(\theta)p(R_1, R_{2,A}, R_{2,D}, F|\theta)p(s|R_1, R_{2,A}, F) \\ &\propto p(\theta)p(R_1, R_2, F|\theta) \end{aligned} \quad (36)$$

$$= p(\theta)p(R_1, R_{2,A}, F|\theta)p(R_{2,D}|R_1, R_{2,A}, F, \theta). \quad (37)$$

As before, the full conditional posterior density for θ corresponds to that from the complete-data problem, using (36). The full conditional posterior density for $R_{2,D}$ is, from (37),

$$p(R_{2,D}|R_1, F, s, \theta) \propto p(R_{2,D}|R_1, R_{2,A}, F, \theta), \quad (38)$$

from which draws require knowledge of only the return likelihood. Also note that the marginal posterior for θ is given by $p(\theta)p(R_1, R_{2,A}, F|\theta)$, the product of the first two factors

⁸Some hedge-fund data services supply data on funds that die after they begin reporting, and such data have been used to compute the survival-bias estimates mentioned earlier. It is not clear, however, that the survival process for such funds is the same as that relevant to the initial reporting of returns.

on the right-hand side of (37). In some problems, where the likelihood $p(R_1, R_{2,A}, F|\theta)$ can be easily specified, such as under the independence assumption in (28), the marginal posterior for θ can be analyzed and computed directly without including a step in which $R_{2,D}$ is drawn.⁹

5. An Illustration: Hedge Funds

Consider a set of four funds ($m_1 = 4$) that have survived through month T , and assume that each fund has ten years of monthly returns history. The funds' ordinary-least-squares (OLS) alphas with respect to the three factors of Fama and French (1993) are 10%, 1%, 0%, and -5% (annualized). Each fund has OLS slopes of 0.6, 0.3, and 0.1 with respect to MKT, SMB, and HML; the sample standard deviation of each fund's residuals is 17% (annualized); and the sample residual correlation for each fund pairing equals zero. The funds' sample estimates for the betas and residual standard deviation are equal, when rounded to the number of digits shown, to the cross sectional averages of those quantities for a sample of equity hedge funds obtained from Hedge Fund Research (HFR).¹⁰

The above cross-section of hedge funds is used in an empirical Bayes fashion to set values for several parameters in the prior distribution of each fund's factor sensitivities. The prior conditional covariance matrix $\Phi_{b|b_0}$ in (9) is set to

$$\Phi_{b|b_0} = C(\hat{b}) - \overline{\hat{\sigma}_i^2(Z'Z)_i^{-1}} = \begin{bmatrix} 0.22 & 0.07 & -0.01 \\ 0.07 & 0.15 & 0.00 \\ -0.01 & 0.00 & 0.23 \end{bmatrix}, \quad (39)$$

where $C(\hat{b})$ is the sample cross-sectional covariance matrix of the funds' OLS \hat{b} 's, $\overline{\hat{\sigma}_i^2(Z'Z)_i^{-1}}$ is the average across funds of the usual estimate for the sampling variance of \hat{b} , and $\hat{\sigma}_i^2$ and $(Z'Z)_i^{-1}$ are based on the observations available for the i -th fund. The prior mean of each fund's b , \bar{b}_0 in (9), is set to the average of the factor sensitivities, but the matrix Φ_{b_0} in (9) is set to a large scalar value times the identity matrix, making the choice of \bar{b}_0 unimportant. Thus, the overall prior belief about the factor sensitivities is that they are, to some degree, similar across assets, but the prior is non-informative with regard to the average value of the sensitivities. In this initial example, the prior for Σ is non-informative. The value s^2 in (43) is set to the average residual variance, but the value of ν in $p(\Sigma)$ in (12) is set to a small

⁹See, for example, Jones and Shanken (2005).

¹⁰The sample consists of the 331 equity funds in the HFR databases of live and dead funds that report returns net of all fees and have at least five years of uninterrupted monthly return history over the period from January 1974 through December 2001.

value ($\nu = m + 2$). An alternative specification with an informative prior for Σ is analyzed later.

For a wide range of prior beliefs about funds' alphas, Table 1 reports the posterior mean of α_i for the fund whose OLS estimate $\hat{\alpha}_i$ equals 10%. Recall that ψ represents prior uncertainty about the cross-sectional mean alpha, α_0 , while σ_α represents prior uncertainty about the deviation of any given fund's α_i from α_0 . At one extreme, when both ψ and σ_α are zero, each fund's α_i is simply constrained to equal the overall unconditional mean, $\bar{\alpha}_0$, which is set to zero for this illustration. When $\psi = 0$, there is no commonality in the prior uncertainty about funds' alphas. This setting corresponds roughly to that in which Baks, Metrick and Wachter (2001) observe that inferences about a surviving fund's parameters are unaffected by conditioning on survival. The correspondence is not complete, however, since in this illustration there is still commonality in prior beliefs about funds' factor betas, whereas Baks et al. assume independence across all fund parameters. (The role of commonality in beliefs about betas is discussed later.) When $\sigma_\alpha = 0$, then all uncertainty about alphas is shared across funds.

The posterior means in Table 1 are reported for two different assumptions about fund survival. In Panel A, there is no conditioning on survival. That is, the four surviving funds are treated as a complete sample that does not reflect the outcome of a survival criterion applied to a larger universe. In Panel B, the posterior mean of the fund's α_i is conditioned on there being four non-surviving funds, so that for every fund included in the observed sample, there was another fund that did not survive. This illustration imposes a simple deterministic survival criterion based on total return: a fund doesn't survive if its cumulative return after the first five years (of the ten-year period) is worse than -5%, which happens to be the worst performance turned in by any of the four surviving funds. Panel C then reports the "survival effect" as the amount by which the value in Panel A exceeds its counterpart in Panel B.

The survival effect is largest when $\sigma_\alpha = 0$ but ψ takes a very large value (denoted here by ∞). An investor with such priors has absolutely no idea what funds' alphas might be, but he is sure that they are all the same. In that case, when there are no non-survivors assumed, the posterior mean of the fund's α_i is 1.5%, which is just the cross-sectional mean of the $\hat{\alpha}_i$'s for the four funds. When four non-survivors are assumed, the posterior mean of α_i drops to -1.7%, so the survival effect is 3.2%. Of course, since $\sigma_\alpha = 0$, these values are also the posterior means for all funds in the sample.

The above case with $\sigma_\alpha = 0$ and $\psi = \infty$ perhaps draws the closest parallel to the empirical studies noted in the introduction, in which a survival bias is estimated by comparing

returns (or abnormal returns) averaged across all surviving funds to those averaged across all surviving and dead funds. To use that value as a correction to be applied in estimating the expected return on a surviving fund, one would presumably first pool that fund's data with others and assume that expected returns are identical across all funds. One would not first estimate a fund's expected return, placing at least some greater weight on the fund's own history than its competitors, and then subtract a value for survival bias as typically computed. The results in Table 1 imply that such a procedure makes too large an adjustment for survival.

The nature of one's prior uncertainty generally plays a large role in inference about α , whether or not survival is considered. In this illustration, and in many settings, the data are simply not sufficiently informative about α so as to make the choice of prior unimportant. At the same time, the form of the prior plays an important role in determining the extent to which conditioning on survival matters for inferences about alphas. The survival effects in Panel C of Table 1 are increasing in ψ . Greater prior uncertainty about α_0 requires greater reliance on the data to infer that cross-sectional mean α_i , and knowing whether the average $\hat{\alpha}_i$ reflects survivorship thus becomes more important. The posterior mean of α_i is increasing in ψ when survivorship is not considered (Panel A) but decreasing in ψ when it is (Panel B). For the smaller values of σ_α and larger values of ψ , the posterior mean switches from positive to negative when moving from the first case to the second. That is, for an investor who seeks a high Sharpe ratio and holds such prior beliefs, a long position in the fund (with $\hat{\alpha} = 10\%$) would improve the best factors-only position when inferences do not account for survival, whereas a short position in the fund becomes attractive once survival is considered.¹¹

Table 2 displays the same analysis for the fund whose $\hat{\alpha}_i$ equals zero. As with the previous fund, the posterior mean of α_i is increasing in ψ in Panel A but typically decreasing in ψ in Panel B. Also note that, although the levels in Panels A and B for this fund are often quite different from those in Table 1, especially as σ_α increases, the survival effects in Panel C are very close to those in Table 1. That is, the survival effect does not depend on $\hat{\alpha}_i$.

The smaller is σ_α , the greater is the degree of shrinkage toward the cross-sectional mean α_0 , which is unknown except when $\psi = 0$. When $\sigma_\alpha = \infty$, the posterior mean of α_i is $\hat{\alpha}_i$ in all cases. With infinite uncertainty about a fund's deviation from the cross-sectional mean α_0 , there is no information provided by either the prior mean for α_0 or by the returns of other funds, living or dead. For the remaining values of σ_α , the survival effects in Panel C

¹¹A portfolio's Sharpe ratio is its expected excess return divided by the standard deviation of its return. See Sharpe (1994).

are decreasing in σ_α . Note that for low values of ψ , the survival effect turns negative for the higher values of σ_α . For the fund with $\hat{\alpha} = 0$, the previously described investor would choose a long position in this (surviving) fund when survival is considered but would choose no position in the same fund when survival is not considered.

The key to the last result, wherein a fund can actually become more attractive when inferences account for survival, lies in uncertainty about factor betas. Recall from (9) that the prior specifies a degree of commonality in the uncertainty about the b_i 's. In this illustration, the prior is non-informative about b_0 , the cross-sectional mean of the b_i 's, but deviations around that common unknown mean have the finite covariance matrix in (39). Thus, there is some degree of shrinkage in posterior means of b_i 's toward b_0 . Accounting for non-surviving funds lowers the posterior mean of b_0 . Funds die because they experience lower returns, and one contributing factor to lower returns is having lower betas on factors that experienced positive returns over the sample period. Thus, accounting for the presence of the non-surviving funds, with their inferred lower betas, also lowers the inferred overall mean mean of b_0 . This effect is strongest when $\psi = 0$ and, therefore, accounting for the poorer performing dead funds cannot lower α_0 . The posterior means of the b_i 's are thus shrunk to a lower b_0 . In general, inferences about alphas and betas tend to be negatively correlated, with lower slopes associated with higher intercepts. Thus, lower posterior means of the b_i 's translate to higher inferred values of α_i 's for the surviving assets than would occur when there is no shrinkage in the b_i 's toward an unknown common mean (i.e., when Φ_{b_0} in (9) is a zero matrix).

Table 3 reports the posterior means of the factor betas for the four surviving funds when $\sigma_\alpha = 10\%$ and $\psi = 0$, a case where the negative survival effects are most pronounced in Tables 1 and 2. Note first that, although the OLS estimates of b_i are identical across the four funds, the posterior means of the b_i 's are increasing in $\hat{\alpha}_i$. This result reflects the shrinkage applied to the α_i 's, coupled with the above-mentioned negative correlation present in the uncertainty about alphas and betas. More importantly, observe that the survival effect in the third row of each panel is positive: about 0.020 for MKT betas, 0.010 for the SMB betas, and 0.025 for the HML betas. That is, the posterior means of the betas are reduced by these amounts when accounting for survival. When applied to the mean returns on the factors over the funds' ten-year sample period (those means are 8.1%, 0.9%, and 6.6% for MKT, SMB, and HML, respectively), these survival effects in the factor betas essentially produce the seemingly reversed survival effects in the posterior mean α_i noted above.

The posterior uncertainty in a fund's alpha is, not surprisingly, highly dependent on the prior. As noted earlier, prior beliefs about alpha are important in samples typically

encountered. Accounting for survival, on the other hand, produces only slight increases in the posterior uncertainty about alpha. Table 4 reports the posterior standard deviation of alpha for the fund whose $\hat{\alpha} = 10\%$. As one would expect, the posterior standard deviation increases in both σ_α and ψ , but the survival effects in Panel C are negligible, generally 10 basis points or less.

6. Style commonality

In the above illustration, recall that the sample correlations between the OLS residuals for all pairs of surviving funds are equal to zero. Suppose instead that the residuals of the surviving funds exhibit positive sample correlations across funds. In other words, suppose the surviving funds exhibit commonality in investment style beyond that captured by the factors in (1). To explore this question, the above calculations are repeated with one change to the observed sample statistics: the sample correlations of the OLS residuals between all pairs of surviving funds are now set to 0.3 instead of zero. All other sample statistics (OLS alphas, betas, and residual standard deviations) remain the same, as does the survival rule (a minimum cumulative return of -5% for the first five years).

Table 5 reports the posterior mean of alpha for the fund whose $\hat{\alpha}_i$ is 10%, so this table is simply the recalculation of Table 1 under the new residual correlation. First note that commonality in style produces somewhat different values in Panel A, as compared to the corresponding values in Table 1. For the positive finite values of σ_α , the values in Table 5 are modestly higher than those in Table 1. This result essentially reflects the fact that differences between sample means become more informative as the correlation among assets increases. With the higher residual correlation in Table 5, more of the observed cross-sectional dispersion in sample means is preserved, and there is less shrinkage toward a central value. This effect does not occur in the first row, since the shrinkage toward a central value is forced to be complete when $\sigma_\alpha = 0$. There, the values in Table 5 for finite positive values of ψ are less than those in Table 1 due to a second effect produced by the style commonality. With higher correlation among assets, the sample becomes less informative about the overall cross-sectional mean, so the prior cross sectional mean (zero in this illustration) is given more weight.

More pronounced is the role of style commonality in survival effects. Note that many of the values in Panel C of Table 5 are substantially higher than those in Table 1, especially with greater prior uncertainty about the overall cross-sectional mean (higher values of ψ).

When the latter uncertainty is very large and σ_α is 5% or less, the survival effects in Table 5 exceed those in Table 1 by at least 200 basis points. The reasoning here is closely related to the second effect in the above discussion of Panel A. Due to the higher residual correlations among the surviving assets, the average of their sample alphas becomes a less informative observation about the overall mean alpha (α_0) of all m assets, including the non-survivors. Thus, the inferred lower alphas of the non-survivors exert a greater effect in depressing the inferred value of α_0 , and the alphas of the surviving assets are thus shrunk toward that lower value. With greater overall uncertainty about α_0 , observed style commonality among the survivors produces larger survival effects.

7. Alpha-related risk

In the analysis thus far, prior beliefs about α and Σ are independent. That is, there is no belief a priori that the intercepts in (1) have anything to do with the covariance matrix of the disturbances. One might instead believe that an asset with a higher absolute alpha is more likely to have higher residual risk, and that two assets whose alphas have the same sign are more likely to have residuals in (1) that are positively correlated. A simple economic motivation for such beliefs is that non-zero alphas represent, at least in part, compensation for additional sources of systematic risk that are not captured by the factors included in (1). (See MacKinlay and Pástor (2000), for example.) Alternatively, one might believe that opportunities for managers to produce positive alphas, if they exist, are more likely to be concentrated within a few investment styles as opposed to distributed across a great many styles. Again, under such a belief, two managers with positive alphas are more likely to have residuals that are positively correlated (i.e., similar styles). Or one might simply believe it unlikely, for whatever reason, that all n ($= m + K$) assets can be deployed to produce Sharpe ratios much higher than can be obtained by combining only the k factors. As shown by Gibbons, Ross, and Shanken (1989),

$$\begin{aligned}\lambda^2 &\equiv S_n^2 - S_k^2 \\ &= \alpha' \Sigma^{-1} \alpha\end{aligned}\tag{40}$$

where S_n^2 is the maximum squared Sharpe ratio for the universe of all n assets and S_k^2 is maximum squared Sharpe ratio that can be obtained by combining the k factors. As demonstrated by MacKinlay (1995) for finite values of m , linking α and Σ in the directions

described above reduces the values of λ^2 (given (40)) that would otherwise occur.¹²

The objective of this section is to investigate whether survival effects are sensitive to the presence of a prior belief in a link between alphas and residual risk. To implement such a link in a probabilistic fashion, the prior for Σ is conditioned on α . That is, the joint prior for all parameters in (20) is replaced by

$$p(\theta) = p(a)p(\Sigma|a)p(\mu)p(\Omega), \quad (41)$$

where $p(\Sigma|a) = p(\Sigma|\alpha)$ as an inverted Wishart density,

$$p(\Sigma|\alpha) \propto |H|^{\frac{\nu}{2}} |\Sigma|^{-\frac{\nu+m+1}{2}} \exp\left\{-\frac{1}{2}\text{tr } H\Sigma^{-1}\right\}, \quad (42)$$

with

$$\text{E}\{\Sigma|\alpha\} = \frac{1}{\nu - m - 1} H = s^2 \left[\left(\frac{\gamma}{1 + \gamma} \right) \frac{1}{(1/m)\text{E}(\alpha'\alpha)} \alpha\alpha' + \left(\frac{1}{1 + \gamma} \right) I_m \right], \quad (43)$$

and the constants s^2 , γ , and ν (the degrees of freedom) must be specified.¹³ (Note from (6) that $\text{E}(\alpha'\alpha) = \text{tr}(\bar{\alpha}_0\bar{\alpha}'_0 + \Phi_\alpha) = \bar{\alpha}'_0\bar{\alpha}_0 + m(\sigma_\alpha^2 + \psi^2)$.) Taking unconditional expectations of (43) gives

$$\frac{1}{m} \text{tr}(\text{E}\{\Sigma\}) = s^2. \quad (44)$$

Thus, s^2 is the average unconditional mean of the residual variance, and, as before, this value can be specified using an empirical Bayes procedure. The degrees of freedom parameter ν controls the strength of the prior, in the sense that, conditional α , the prior supplies the same information about Σ as does a sample of size ν with residual covariance matrix given by (43).

In the above specification, the conditional prior mean of Σ in (43) is the sum of matrices proportional to $\alpha\alpha'$ and I_m . MacKinlay and Pástor (2000) propose a similar restriction on Σ in maximum-likelihood estimation. The higher the value of γ (≥ 0), the greater is the prior expected role of α in the residual covariance matrix. An asset's residual variance becomes more closely linked to its absolute alpha, and the correlation between the residual returns of

¹²Note also that the portfolio with the maximum absolute Sharpe ratio also attains the maximum absolute information ratio, where the information ratio is a portfolio's alpha divided by its residual standard deviation. Moreover, it is easily verified that the maximum squared information ratio is given by

$$\max_w \frac{(w'\alpha)^2}{w'\Sigma w} = \alpha'\Sigma^{-1}\alpha = \lambda^2.$$

¹³Equivalently, the prior for Σ^{-1} given α is Wishart with parameter matrix H^{-1} and degrees of freedom equal to ν .

any two assets i and j moves closer to $(\alpha_i\alpha_j)/\sqrt{\alpha_i^2\alpha_j^2} = \pm 1$, where the sign is that of $\alpha_i\alpha_j$. Further insight into the magnitude of γ can be obtained by first observing that

$$\lambda^2|\alpha \sim (\alpha'H^{-1}\alpha)\chi_\nu^2, \quad (45)$$

using (40) and properties of the Wishart distribution. The inverse of H is given by

$$H^{-1} = \frac{1 + \gamma}{(\nu - m - 1)s^2} \left[I_m - \frac{(m\gamma)/E(\alpha'\alpha)}{1 + (m\gamma\alpha'\alpha)/E(\alpha'\alpha)} \alpha\alpha' \right], \quad (46)$$

so

$$\alpha'H^{-1}\alpha = \frac{1 + \gamma}{(\nu - m - 1)s^2} \left(\frac{\alpha'\alpha}{1 + (m\gamma\alpha'\alpha)/E(\alpha'\alpha)} \right). \quad (47)$$

Combining (45) and (47) gives

$$\lambda^2|\alpha \sim \left[\left(\frac{1}{\nu - m - 1} \chi_\nu^2 \right) \left(\frac{\alpha'\alpha}{s^2} \right) \right] \times \left(\frac{1 + \gamma}{1 + (m\gamma\alpha'\alpha)/E(\alpha'\alpha)} \right). \quad (48)$$

The first factor on the right-hand side of (48), in brackets, is the value of λ^2 with $\gamma = 0$, with no prior link between α and Σ . When $\gamma > 0$, the second factor is less than unity and produces a downward adjustment. For $\alpha'\alpha$ equal to its prior mean, that adjustment factor is equal to $(1 + \gamma)/(1 + m\gamma)$.

A prior association between α and Σ is also a feature of the Bayesian settings in Pástor (2000) and Pástor and Stambaugh (1999, 2001, 2002a, 2002b). In those studies, a non-informative marginal prior for Σ is specified. Then, in the conditional prior for α given Σ , the variance of α is proportional to Σ . Such a specification also associates alpha and residual variance in the directions described earlier: a larger diagonal value in Σ is more likely to be associated with a larger absolute value of the corresponding entry in α , and a larger positive (negative) residual covariance between two assets is more likely to be associated with entries in α of the same (opposite) signs. In the alternative specification above, a marginal prior is specified for α instead of Σ so that beliefs about commonality in the elements of α can be expressed as in (6).

Table 6 reports results in which the dataset is identical to that used in Table 5, with sample residual correlations of 0.3, but where the non-informative prior for Σ is replaced by two different versions of the conditional prior distribution in (42). In the first version γ is set to zero, introducing no prior link between α and σ , whereas in the second version γ is set to 0.025. With $m = 8$ and $\gamma = 0.025$, as in the case illustrated here, the second factor on the right-hand side of (48) equals approximately 0.85 when $\alpha'\alpha = E(\alpha'\alpha)$. For $\alpha'\alpha$ equal to twice its prior mean, that downward adjustment factor drops to 0.73. Under

both values for γ , the degrees of freedom value ν in (42) is set so $\nu - m = 100$, where m is the total number of assets. Thus, the conditional prior for Σ introduces about as much information as a hypothetical sample of 100 months. (Recall that the actual sample size is 120 months.) To preclude differences in m from playing any role in this analysis (since m enters (42) and (43)), the total number of assets is held at $m = 8$ throughout the calculations in Table 6, even in Panel A where inferences are based only on the four survivors and do not condition on survival. Calculations in the latter case simply use the degenerate survival rule, $p(s|R_1, R_2, F) = 1$ for all R_1, R_2 , and F .

The first three columns in Table 6 display, for a basis of comparison, the results with no prior association between α and Σ , i.e., $\gamma = 0$. The posterior means of alpha and the survival effects in this case lie between their corresponding values in Tables 1 and 5. The prior mean of Σ is proportional to the identity matrix when $\gamma = 0$. That is also the prior mean in Table 5, where $\gamma = 0$ as well, but there $\nu - m = 2$ so the posterior for Σ depends almost entirely on the sample residual covariance matrix, in which the residual correlation is 0.3. With $\nu - m = 100$, as in Table 6, the residual covariance matrix is shrunk away from that sample estimate and toward the outcome in Table 1, in which the sample covariance matrix is proportional to the identity.

The results in Table 6 when prior beliefs include an association between α and Σ , are shown in the three rightmost columns of Table 6. The intuition for the outcomes in this case is more elusive, because the posterior means reflect a combination of multiple influences. The sample $\hat{\alpha}$'s have little impact on the posterior mean of Σ , essentially because the sample residual covariance matrix contains more precise information about Σ . On the other hand, that sample covariance matrix does exert a substantial influence on the posterior distribution of α , through the prior link between α and Σ . In Panel A, with no survival effect, the posterior means of the fund whose sample alpha is 10% are often larger, by substantial amounts, than the corresponding entries with $\gamma = 0$. In contrast, when conditioned on survival in Panel B, the fund's posterior mean alphas are often substantially lower than with $\gamma = 0$. In essence, the link between α and Σ often, in this illustration, tends to increase the absolute value of the inferred alphas relative to what they would have been otherwise, but the sign is more likely to be determined by the other sources of information in the prior and the data. Conditioning on survival flips some of the inferred alphas (among all eight funds) from positive to negative, and then the link with Σ often amplifies their absolute values. The combined result is that Panel C contains the largest survival effects yet observed in this scenario with four surviving funds.

8. Conclusions

In general, conditioning on an asset's having survived affects inferences about that asset's return distribution. Even when the survival process is known, inferences about the surviving asset's expected return or alpha depend critically on the extent to which one's prior uncertainty about parameter values is common across assets. If a priori knowledge of one fund's alpha provides information about another fund's alpha, before conditioning on returns data or survival outcomes, then conditioning on survival becomes more important in inferring the alpha of a surviving fund. So, for example, if alpha is entertained as a measure of managerial skill, then such beliefs would translate roughly as a view that discovery of one skilled manager probably points to others. Complete certainty that all assets have the same alpha coupled with infinite uncertainty about what that alpha might be represents an extreme version of commonality in prior uncertainty, and the effects of conditioning on survival in that case are typically the strongest. Positive correlations among the observed non-factor returns on assets, indicating what is termed here as style commonality, also lead to stronger effects from conditioning on survival.

At another extreme are prior beliefs in which parameter uncertainty is independent across assets. When coupled with an assumption of independence in factor-adjusted returns (residuals), such prior beliefs lead to the result that conditional inference about survivors is unaffected by conditioning on their survival, even though the average returns or sample alphas of such funds are nevertheless biased in classical sampling terms.

What is the appropriate degree of commonality in one's prior uncertainty about assets' expected returns or alphas? The answer no doubt differs across individuals, but it probably depends on the application as well. A fair amount of commonality would seem reasonable if the assets are, say, aggregate equity portfolios in different developed countries. There, knowing the true value of the equity premium in one country could well be useful prior information about the premium in another country, given a view that preferences and production technologies are similar across countries or that capital markets are integrated. If instead the assets are mutual funds, then the complete absence of commonality in uncertainty about alphas, as in Baks, Metrick, and Wachter (2001), might not be unreasonable, especially if one views alpha as a measure of skill. It seems plausible that, for at least some investors, learning one manager is skilled might not lead them to revise their beliefs about the skill of another manager. For other investors, though, such information might cause them to look upon another manager more favorably.

Usually, but not always, conditioning on survival lowers the inferred alphas for assets that passed a survival test failed by assets with lower returns. Exceptions can occur when there is little if any commonality in the prior uncertainty about alphas but there is commonality in one's prior uncertainty about assets' factor sensitivities (betas). Conditioning on survival can then lower the inferred betas for surviving funds and thereby raise their alphas. This effect leads to the paradoxical result that an investor already holding the best factors-only position could be interested in establishing a new long position in a surviving asset only when conditioning on that asset's having survived.

As noted at the outset, in order to simplify the analysis and sharpen the focus on other key aspects of the inference problem, the survival process in the illustrations presented here is taken as a known deterministic function of an asset's return path. Clearly, interesting directions for further research would be to relax the deterministic nature of survival and make the process stochastic (conditional on the price path), and to allow prior uncertainty about the nature of the survival process.

Appendix

Let F_T denote the $T \times k$ matrix containing realizations of the factors corresponding to the T periods in R , and define $Z = (\iota_T \ F_T)$, where ι_T denotes a T -vector of ones. For those T observations, the regression model in (1) can be written as

$$R = ZA + U, \quad \text{vec}(U) \sim N(0, \Sigma \otimes I_T), \quad (\text{A.1})$$

where $U = (u_1, \dots, u_T)'$. Define the statistics $\hat{A} = (Z'Z)^{-1}Z'Y$ and $\hat{\Sigma} = (R - Z\hat{A})'(R - Z\hat{A})/T$. Also let $\hat{a} = \text{vec}(\hat{A})$. Let F_B denote the $T_B \times k$ matrix containing the earlier observations of the benchmark factors, so that the total sample of $T_L = T_B + T$ observations of the factors is contained in the matrix

$$F = \begin{bmatrix} F_B \\ F_T \end{bmatrix}. \quad (\text{A.2})$$

Define $\hat{\mu} = F'\iota_{T_L}/T_L$, and $\hat{\Omega} = (F - \iota_{T_L}\hat{\mu})'(F - \iota_{T_L}\hat{\mu})/T_L$.

Recall that the last m_2 columns of R are unobserved, and that θ denotes the set of all unknown parameters. The complete-sample likelihood function, $p(R, F|\theta)$, appears as the second factor on the right-hand side of (26). This likelihood can be factored as

$$p(R, F|\theta) = p(R|\theta, F_T) p(F|\theta), \quad (\text{A.3})$$

where

$$\begin{aligned} p(R|\theta, F_T) &\propto |\Sigma|^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2} \text{tr} (R - ZA)'(R - ZA)\Sigma^{-1} \right\} \\ &\propto |\Sigma|^{-\frac{T}{2}} \exp \left\{ -\frac{T}{2} \text{tr} \hat{\Sigma}\Sigma^{-1} - \frac{1}{2} \text{tr} (A - \hat{A})'Z'Z(A - \hat{A})\Sigma^{-1} \right\} \\ &\propto |\Sigma|^{-\frac{T}{2}} \exp \left\{ -\frac{T}{2} \text{tr} \hat{\Sigma}\Sigma^{-1} - \frac{1}{2} (a - \hat{a})'(\Sigma^{-1} \otimes Z'Z)(a - \hat{a}) \right\} \end{aligned} \quad (\text{A.4})$$

and

$$\begin{aligned} p(F|\theta) &\propto |\Omega|^{-\frac{T_L}{2}} \exp \left\{ -\frac{1}{2} \text{tr} (F - \iota_{T_L}\hat{\mu})'(F - \iota_{T_L}\hat{\mu})\Omega^{-1} \right\} \\ &\propto |\Omega|^{-\frac{T_L}{2}} \exp \left\{ -\frac{T_L}{2} \text{tr} \hat{\Omega}\Omega^{-1} - \frac{T_L}{2} \text{tr} (\mu - \hat{\mu})(\mu - \hat{\mu})'\Omega^{-1} \right\}. \end{aligned} \quad (\text{A.5})$$

Note that in the i.i.d. stochastic setting assumed here, $p(R|\theta, F_T) = p(R|\theta, F)$.

The conditional density $p(R_2|R_1, F, \theta)$, the first factor on the right-hand side of (27), is a product of multivariate normal densities, with

$$r_{2,t}|R_1, F, \theta \sim N(E_{2,t|\cdot}, V_{22|\cdot}), \quad (\text{A.6})$$

$$E_{2,t|\cdot} = \alpha_2 + B_2 f_t + \Sigma_{21} \Sigma_{11}^{-1} (r_{1,t} - \alpha_1 - B_1 f_t), \quad (\text{A.7})$$

$$V_{22|\cdot} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}, \quad (\text{A.8})$$

and the partitions are defined as

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad \text{and} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}. \quad (\text{A.9})$$

Posterior distributions are computed using an MCMC approach. At each iteration of the chain, a new draw of the non-observed returns R_2 is obtained, using (27). In this implementation of the approach, s (survival) is assumed to be a non-random function $s(R_1, R_2, F)$, so the density $p(s|R_1, R_2, F)$, the second factor on the right-hand side of (27), is an indicator function. That is, $p(s|R_1, R_2, F) = 1$ if $s(R_1, R_2, F)$ equals the observed s , whose first m_1 elements equal 1 and whose second m_2 elements equal 0; $p(s|R_1, R_2, F) = 0$ otherwise. To obtain a draw of R_2 from (27), a candidate value can be drawn from the density $p(R_2|R_1, F, \theta)$ in (A.6) and then retained as the draw from $p(R_2|R_1, F, s, \theta)$ if m_2 assets with those returns would not have survived. If, given R_1 and F , a candidate asset's non-survival depends only on its own hypothetical return history (and not on other non-surviving assets), it is often more efficient to draw the non-observed returns one asset at a time, conditional on the m_1 observed return histories (R_1) as well as the histories for the $m' (< m_2)$ non-observed assets that have already been drawn (the first m' columns of R_2). To condition in that fashion, the candidate returns for the next missing asset are drawn as in (A.6), except with the partitioning redefined so that the first block contains all of the $m_1 + m'$ assets on which the current draw is being conditioned.

The full conditional posterior for a is obtained by taking the product of the prior densities in (42) and (17) and the likelihood in (A.4). First note that, when viewed as a function of a , the right-hand side of (42) can be written as

$$\begin{aligned} g(a) &= |H|^{\frac{\nu}{2}} |\Sigma|^{\frac{\nu+m+1}{2}} \exp\left\{-\frac{1}{2} \text{tr} H \Sigma^{-1}\right\} \\ &= c(1 + h_2 \alpha' \alpha)^{\frac{\nu}{2}} \exp\left\{-\frac{h_1 h_2}{2} \alpha' \Sigma^{-1} \alpha\right\}, \end{aligned} \quad (\text{A.10})$$

with c a constant not involving a , using the relation $|H| = h_1^m (1 + h_2 \alpha' \alpha)$ after writing (43) more compactly as $H = h_1 (h_2 \alpha \alpha' + I_m)$. Multiplying (17) and (A.10) by the likelihood in the third line of (A.4) then gives

$$\begin{aligned}
p(a|\cdot) &\propto \exp\left\{-\frac{1}{2}\left[(a-\bar{a})'\Phi_a^{-1}(a-\bar{a})\right]\right\} \\
&\quad \times (1+h_2\alpha'\alpha)^{\frac{\nu}{2}} \exp\left\{-\frac{h_1h_2}{2}\alpha'\Sigma^{-1}\alpha\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}(a-\hat{a})'(\Sigma^{-1}\otimes Z'Z)(a-\hat{a})\right\}. \tag{A.11}
\end{aligned}$$

Rather than draw a directly from (A.11), a Metropolis-Hastings procedure is applied.¹⁴ Omitting the first factor in the second line of (A.11) gives a normal proposal density $p^\dagger(a|\cdot)$, where the covariance matrix can be scaled if necessary. That is, the proposal density is given by

$$a|\cdot \sim N(\tilde{a}^\dagger, \zeta\tilde{V}_a^\dagger), \tag{A.12}$$

where

$$\tilde{V}_a^\dagger = \left(\Phi_a^{-1} + h_1h_2D\Sigma^{-1}D' + \Sigma^{-1}\otimes Z'Z\right)^{-1}, \tag{A.13}$$

$$\tilde{a}^\dagger = \tilde{V}_a^\dagger\left(\Phi_a^{-1}\bar{a} + (\Sigma^{-1}\otimes Z'Z)\hat{a}\right), \tag{A.14}$$

$D = I_m \otimes d$, d is a $(k+1)$ -element column vector with 1 as the first element and 0 elsewhere, and ζ is positive scalar that can be tuned to improve acceptance rates and tail coverage. Let a_n denote the current value in the chain of draws, and let a_c denote a candidate value drawn from $p^\dagger(a|\cdot)$. Compute the ratio

$$\begin{aligned}
\kappa &= \frac{p(a_c|\cdot)p^\dagger(a_n|\cdot)}{p(a_n|\cdot)p^\dagger(a_c|\cdot)} \\
&= \frac{(1+h_2\alpha'_c\alpha_c)^{\frac{\nu}{2}} \exp\left\{-\frac{\zeta-1}{2\zeta}(a_c-\tilde{a}^\dagger)'\tilde{V}_a^{\dagger-1}(a_c-\tilde{a}^\dagger)\right\}}{(1+h_2\alpha'_n\alpha_n)^{\frac{\nu}{2}} \exp\left\{-\frac{\zeta-1}{2\zeta}(a_n-\tilde{a}^\dagger)'\tilde{V}_a^{\dagger-1}(a_n-\tilde{a}^\dagger)\right\}}, \tag{A.15}
\end{aligned}$$

where α_n and α_c denote the current and candidate values, respectively. With probability $\min(1, \kappa)$, replace a_n with a_c as the current value of a in the Markov chain; otherwise retain a_n as the current value.

The full conditional posterior for Σ is obtained by taking the product of the prior density in (42) and the likelihood as expressed in the second line of (A.4),

$$p(\Sigma|\cdot) \propto |\Sigma|^{-\frac{T+\nu+m+1}{2}} \exp\left\{-\frac{1}{2}\text{tr}\left[H + T\hat{\Sigma} + (A-\hat{A})'Z'Z(A-\hat{A})\right]\right\}, \tag{A.16}$$

which is an inverse Wishart density, so

$$\Sigma^{-1}|\cdot \sim W\left(T+\nu, \left[H + T\hat{\Sigma} + (A-\hat{A})'Z'Z(A-\hat{A})\right]^{-1}\right), \tag{A.17}$$

¹⁴See Chib and Greenberg (1995) for a discussion of the Metropolis-Hastings algorithm.

where H is evaluated at the most recent value of α , using (43).

The full conditional for μ is obtained by taking the product of the prior density in (15) and the likelihood in (A.5),

$$p(\mu|\cdot) \propto \exp \left\{ -\frac{1}{2} \left[(\mu - \bar{\mu})' \Phi_{\mu}^{-1} (\mu - \bar{\mu}) + T_L (\mu - \hat{\mu})' \Omega^{-1} (\mu - \hat{\mu}) \right] \right\}, \quad (\text{A.18})$$

so

$$\mu|\cdot \sim (\tilde{\mu}_{|\cdot}, \tilde{V}_{\mu|\cdot}), \quad (\text{A.19})$$

where

$$\tilde{V}_{\mu|\cdot} = [\Phi_{\mu}^{-1} + T_L \Omega^{-1}]^{-1}. \quad (\text{A.20})$$

and

$$\tilde{\mu}_{|\cdot} = \tilde{V}_{\mu|\cdot} [\Phi_{\mu}^{-1} \bar{\mu} + T_L \Omega^{-1} \hat{E}_2]. \quad (\text{A.21})$$

The full conditional for Ω is obtained by taking the product of the prior density in (16) and the likelihood in (A.5),

$$p(\Omega|\cdot) \propto |\Omega|^{-\frac{T_L+k+1}{2}} \exp \left\{ -\frac{T_L}{2} [\hat{\Omega} + (\mu - \hat{\mu})(\mu - \hat{\mu})'] \Omega^{-1} \right\}, \quad (\text{A.22})$$

so

$$\Omega^{-1}|\cdot \sim W \left(T_L, \frac{1}{T_L} [\hat{\Omega} + (\mu - \hat{\mu})(\mu - \hat{\mu})']^{-1} \right). \quad (\text{A.23})$$

The predictive mean and covariance matrix of returns for the upcoming month are given here for the entire returns vector r_{T+1} , although interest is likely to center on only the subset corresponding to the moments of $r_{1,T+1}$, the returns for the upcoming month on the m_1 observed assets. At each iteration of the chain, use (4) and (5) along with the current values of a , (reshaped as α and B), Σ , μ , and Ω to compute the corresponding values of E and V . The posterior mean and variance of E , $E(E|R_1, F, s)$ and $\text{Var}(E|R_1, F, s)$, are computed as the mean and covariance matrix of the MCMC draws of E . The posterior mean of V , $E(V|R_1, F, s)$, is computed as the mean of the MCMC draws of V . The predictive mean of r_{T+1} is given by

$$E(r_{T+1}|R_1, F, s) = E(E|R_1, F, s). \quad (\text{A.24})$$

Using variance decomposition, the predictive return variance is given by

$$\begin{aligned} \text{Var}(r_{T+1}|R_1, F, s) &= E\{\text{Var}(r_{T+1}|R_1, F, s, \theta)|R_1, F, s\} \\ &\quad + \text{Var}\{E(r_{T+1}|R_1, F, s, \theta)|R_1, F, s\} \end{aligned} \quad (\text{A.25})$$

$$= E(V|R_1, F, s) + \text{Var}(E|R_1, F, s). \quad (\text{A.26})$$

Thus, the predictive moments are obtained directly from the MCMC procedure.

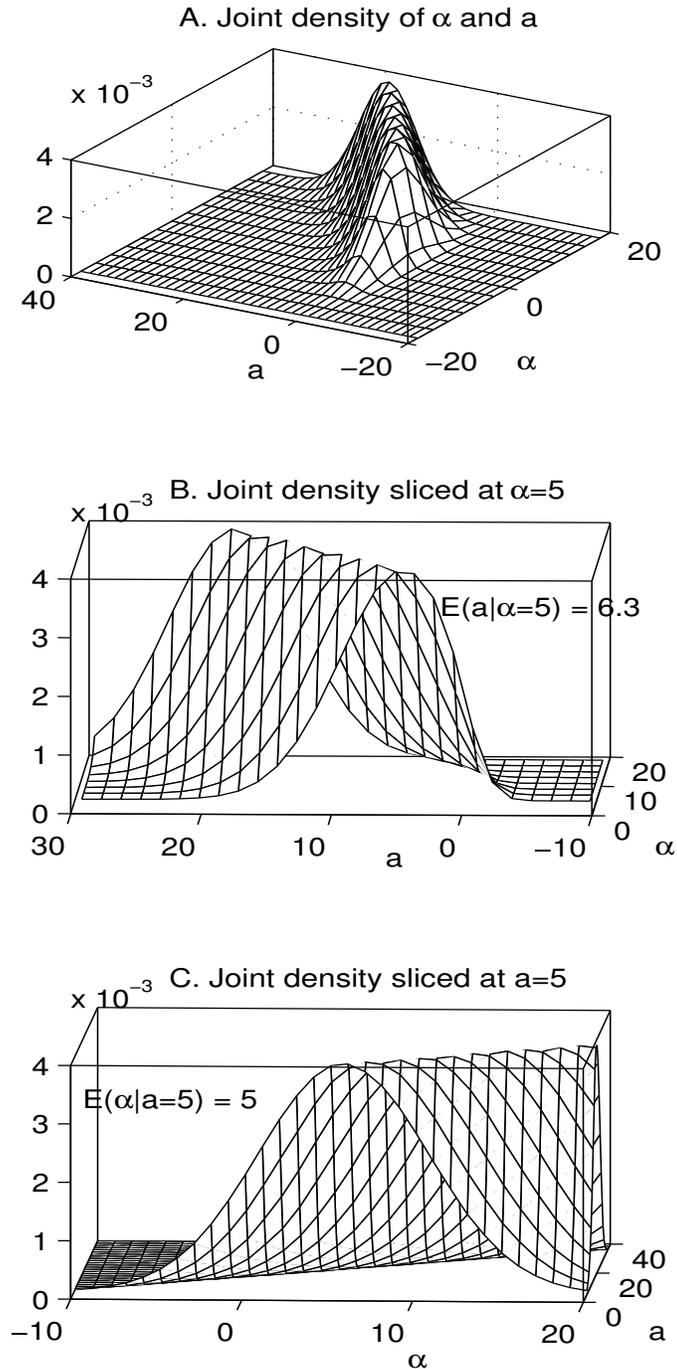


Figure 1. Joint density of an asset's unconditional expected return (α) and its sample mean (a) conditioned on surviving a fixed-return threshold. Both α and a are expressed on an annualized basis. Their joint realizations are obtained by drawing α from a uniform distribution on the interval $[-20\%, 20\%]$, generating a 120-month return history using a mean of α and a 15% volatility, then retaining the joint realization of (α, a) only if that sample's cumulative return is no worse than -10%. Panel B slices the joint density at $\alpha = 5\%$, revealing the conditional density of a , and $E(a|\alpha = 5.0\%) = 6.3\%$. Panel C slices the joint density at $a = 5\%$, revealing the conditional density of α , and $E(\alpha|a = 5.0\%) = 5.0\%$.

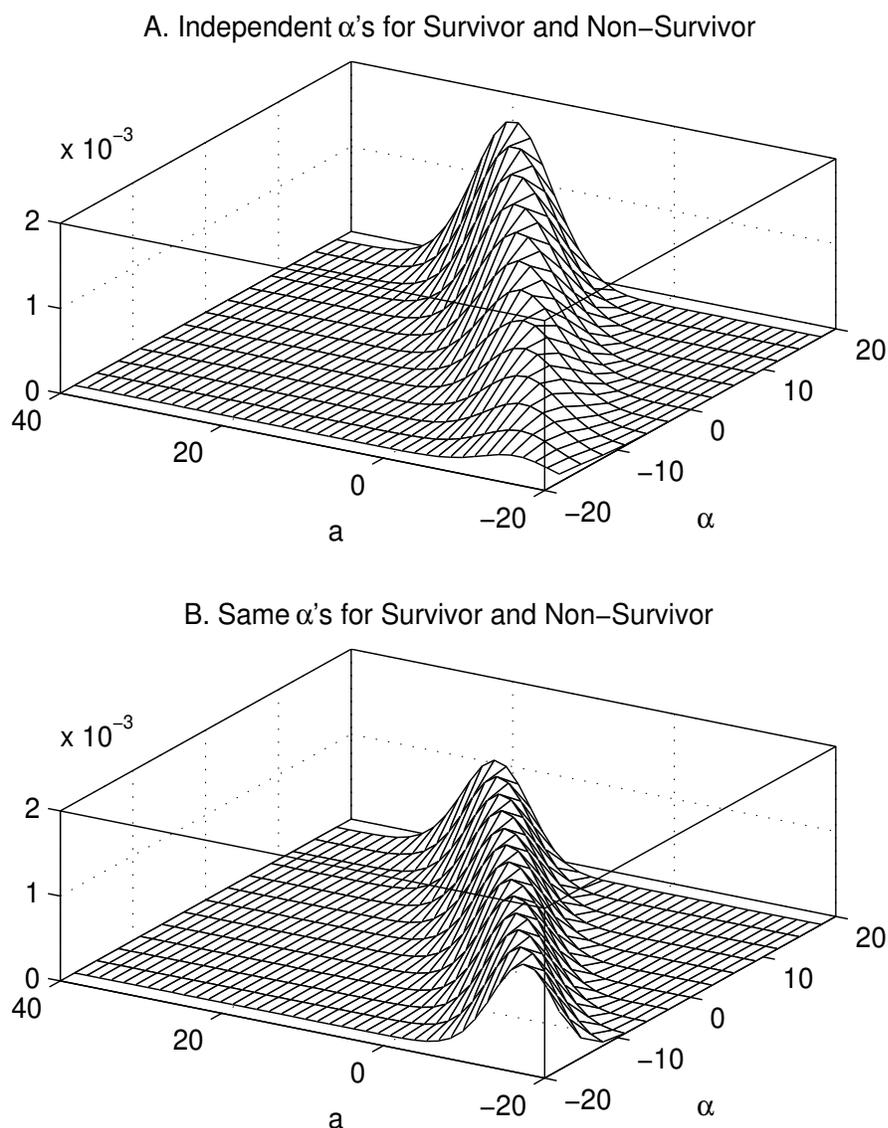


Figure 2. Joint density of an asset's unconditional expected return (α) and its sample mean (a) conditioned on having outperformed another asset. Both α and a are expressed on an annualized basis. Their joint realizations are obtained by drawing α from a uniform distribution on the interval $[-20\%, 20\%]$, generating a 120-month return history using a mean of α and a 15% volatility, then retaining the joint realization of (α, a) only if that sample's cumulative return over the sample period exceeds that of an independently drawn sample for of another "non-surviving" asset with the same 15% volatility. In Panel A, the non-surviving asset's α is drawn from the uniform distribution on $[-20\%, 20\%]$, independently of the surviving fund's α . In Panel B, the non-surviving fund's α is equal to the surviving fund's α .

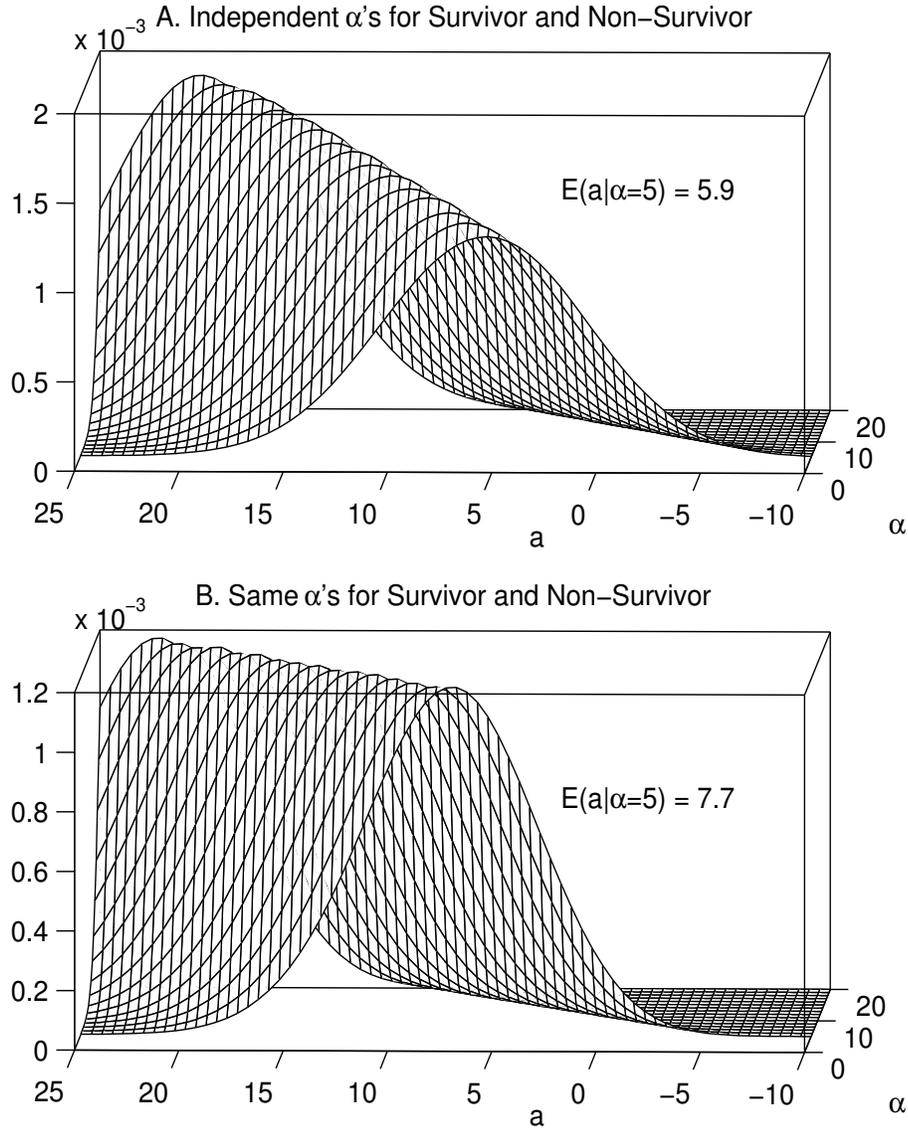


Figure 3. Joint densities in Figure 2 sliced at $\alpha = 5\%$. Each slice reveals the density of the surviving fund's sample mean (a) conditional on having a true unconditional expected return (α) equal to 5%. Both α and a are expressed on an annualized basis. Their joint realizations are obtained by drawing α from a uniform distribution on the interval $[-20\%, 20\%]$, generating a 120-month return history using a mean of α and a 15% volatility, then retaining the joint realization of (α, a) only if that sample's cumulative return over the sample period exceeds that of an independently drawn sample for of another "non-surviving" asset with the same 15% volatility. In Panel A, the non-surviving asset's α is drawn from the uniform distribution on $[-20\%, 20\%]$, independently of the surviving fund's α . In that case, $E(a|\alpha = 5.0\%) = 5.9\%$. In Panel B, the non-surviving fund's α is equal to the surviving fund's α . In that case, $E(a|\alpha = 5.0\%) = 7.7\%$.

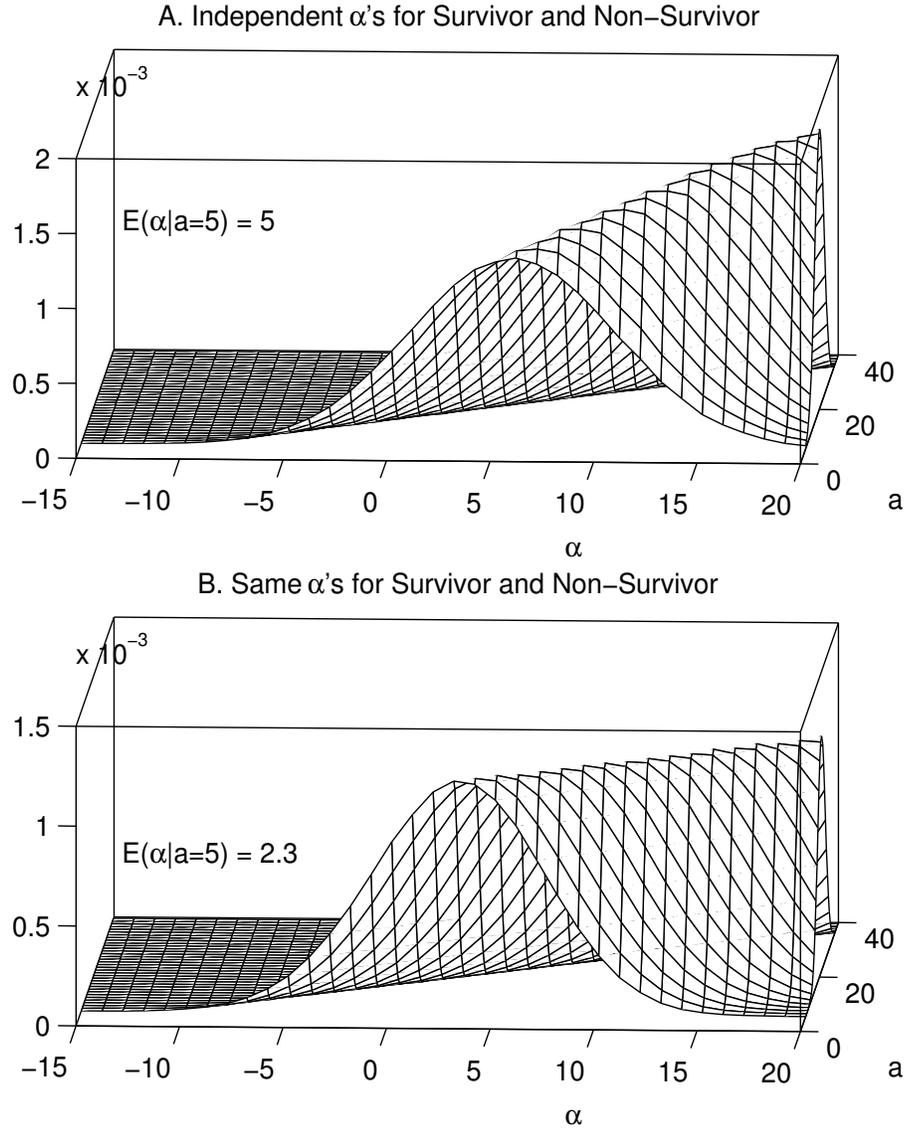


Figure 4. Joint densities in Figure 2 sliced at $a = 5\%$. Each slice reveals the density of the surviving fund's true unconditional expected return (α) conditional on having a sample mean (a) equal to 5%. Both α and a are expressed on an annualized basis. Their joint realizations are obtained by drawing α from a uniform distribution on the interval $[-20\%, 20\%]$, generating a 120-month return history using a mean of α and a 15% volatility, then retaining the joint realization of (α, a) only if that sample's cumulative return over the sample period exceeds that of an independently drawn sample for of another "non-surviving" asset with the same 15% volatility. In Panel A, the non-surviving asset's α is drawn from the uniform distribution on $[-20\%, 20\%]$, independently of the surviving fund's α . In that case, $E(\alpha|a = 5.0\%) = 5.0\%$. In Panel B, the non-surviving fund's α is equal to the surviving fund's α . In that case, $E(\alpha|a = 5.0\%) = 2.3\%$.

Table 1
Posterior Mean of Alpha for a Surviving Fund
Whose Sample Alpha is 10 Percent

The fund is one of four surviving funds, each of which has a ten-year return history. The other three surviving funds have sample OLS alphas of 1 percent, 0 percent, and -5 percent per annum. All four surviving funds have (i) sample OLS betas of 0.6, 0.3, and 0.1 with respect to the three Fama-French factors MKT, SMB, and HML, (ii) sample residual standard deviations of 17 percent per annum, and (iii) OLS residuals that are uncorrelated across funds. The true alpha for any fund i (surviving or not) is denoted by α_i and has a prior distribution of the form

$$\alpha_i|\alpha_0 \sim N(\alpha_0, \sigma_\alpha^2), \quad \alpha_0 \sim N(0, \psi^2).$$

The assumed number of non-surviving funds, whose returns are unobserved, is zero in Panel A and four in Panel B. All non-surviving funds are assumed to have a cumulative return for the first five years below -5%, the lowest initial five-year return of the surviving funds in this sample. The values in the table are expressed in percent per year.

σ_α (%)	ψ (%)					
	0	1	2	5	10	∞
A. Four survivors, zero non-survivors						
0	0.0	0.2	0.5	1.1	1.4	1.5
2	1.1	1.2	1.5	2.0	2.3	2.4
5	4.3	4.3	4.5	4.8	5.0	5.1
10	7.5	7.5	7.5	7.6	7.8	7.9
∞	10.0	10.0	10.0	10.0	10.0	10.0
B. Four survivors, four non-survivors						
0	-0.0	-0.2	-0.6	-1.3	-1.6	-1.7
2	1.2	1.0	0.6	-0.2	-0.5	-0.6
5	4.6	4.6	4.2	3.4	3.0	2.7
10	7.9	7.9	7.7	7.1	6.6	6.2
∞	10.0	10.0	10.0	10.0	10.0	10.0
C. Survival effect: Panel A minus Panel B						
0	0.0	0.3	1.1	2.4	2.9	3.2
2	-0.1	0.2	0.9	2.3	2.8	3.1
5	-0.4	-0.2	0.3	1.4	2.1	2.4
10	-0.4	-0.3	-0.1	0.5	1.2	1.6
∞	0.0	0.0	0.0	0.0	0.0	0.0

Table 2
Posterior Mean of Alpha for a Surviving Fund
Whose Sample Alpha is Zero

The fund is one of four surviving funds, each of which has a ten-year return history. The other three surviving funds have sample OLS alphas of 10 percent, 1 percent, and -5 percent per annum. All four surviving funds have (i) sample OLS betas of 0.6, 0.3, and 0.1 with respect to the three Fama-French factors MKT, SMB, and HML, (ii) sample residual standard deviations of 17 percent per annum, and (iii) OLS residuals that are uncorrelated across funds. The true alpha for any fund i (surviving or not) is denoted by α_i and has a prior distribution of the form

$$\alpha_i|\alpha_0 \sim N(\alpha_0, \sigma_\alpha^2), \quad \alpha_0 \sim N(0, \psi^2).$$

The assumed number of non-surviving funds, whose returns are unobserved, is zero in Panel A and four in Panel B. All non-surviving funds are assumed to have a cumulative return for the first five years below -5%, the lowest initial five-year return of the surviving funds in this sample. The values in the table are expressed in percent per year.

σ_α (%)	ψ (%)					
	0	1	2	5	10	∞
A. Four survivors, zero non-survivors						
0	0.0	0.2	0.5	1.1	1.4	1.5
2	0.0	0.1	0.4	1.0	1.2	1.3
5	0.0	0.1	0.2	0.5	0.7	0.9
10	0.0	0.0	0.0	0.2	0.3	0.4
∞	0.0	0.0	0.0	0.0	0.0	0.0
B. Four survivors, four non-survivors						
0	0.0	-0.2	-0.6	-1.3	-1.6	-1.7
2	0.1	-0.1	-0.5	-1.3	-1.6	-1.7
5	0.2	0.1	-0.2	-1.0	-1.4	-1.5
10	0.2	0.3	0.0	-0.4	-0.9	-1.2
∞	0.0	0.0	0.0	0.0	0.0	0.0
C. Survival effect: Panel A minus Panel B						
0	0.0	0.3	1.1	2.4	2.9	3.2
2	-0.1	0.3	0.9	2.3	2.8	3.0
5	-0.2	-0.1	0.4	1.5	2.1	2.4
10	-0.3	-0.3	-0.0	0.6	1.2	1.5
∞	0.0	0.0	0.0	0.0	0.0	0.0

Table 3
Posterior Means of Factor Betas
 $(\sigma_\alpha = 10\%, \psi = 0)$

The table displays posterior means of factor betas for four surviving funds, each with a ten-year return history. Each surviving fund has a different OLS alpha ($\hat{\alpha}$) but all four funds have (i) sample OLS betas of 0.6, 0.3, and 0.1 with respect to the three Fama-French factors MKT, SMB, and HML, (ii) sample residual standard deviations of 17 percent per annum, and (iii) OLS residuals that are uncorrelated across funds. The true alpha for any fund i (surviving or not) is denoted by α_i and has a prior distribution of the form

$$\alpha_i | \alpha_0 \sim N(\alpha_0, \sigma_\alpha^2), \quad \alpha_0 \sim N(0, \psi^2).$$

The assumed number of non-surviving funds, whose returns are unobserved, is shown in the first column. All non-surviving funds are assumed to have a cumulative return for the first five years below -5%, the lowest initial five-year return of the surviving funds in this sample. The $\hat{\alpha}$'s are expressed in percent per year.

non- survivors	$\hat{\alpha}$			
	10%	1%	0%	-5%
A. MKT betas				
0	0.616	0.602	0.600	0.592
4	0.597	0.581	0.580	0.572
diff.	0.019	0.021	0.020	0.021
B. SMB betas				
0	0.307	0.301	0.300	0.297
4	0.298	0.292	0.292	0.288
diff.	0.009	0.009	0.009	0.009
C. HML betas				
0	0.119	0.102	0.100	0.092
4	0.095	0.078	0.077	0.066
diff.	0.024	0.025	0.024	0.025

Table 4
Posterior Standard Deviation of Alpha for a Surviving Fund
Whose Sample Alpha is 10 Percent

The fund is one of four surviving funds, each of which has a ten-year return history. The other three surviving funds have sample OLS alphas of 1 percent, 0 percent, and -5 percent per annum. All four surviving funds have (i) sample OLS betas of 0.6, 0.3, and 0.1 with respect to the three Fama-French factors MKT, SMB, and HML, (ii) sample residual standard deviations of 17 percent per annum, and (iii) OLS residuals that are uncorrelated across funds. The true alpha for any fund i (surviving or not) is denoted by α_i and has a prior distribution of the form

$$\alpha_i|\alpha_0 \sim N(\alpha_0, \sigma_\alpha^2), \quad \alpha_0 \sim N(0, \psi^2).$$

The assumed number of non-surviving funds, whose returns are unobserved, is zero in Panel A and four in Panel B. All non-surviving funds are assumed to have a cumulative return for the first five years below -5%, the lowest initial five-year return of the surviving funds in this sample. The values in the table are expressed in percent per year.

σ_α (%)	ψ (%)					
	0	1	2	5	10	∞
A. Four survivors, zero non-survivors						
0	0.0	0.9	1.7	2.5	2.8	2.9
2	1.9	2.1	2.4	3.0	3.3	3.4
5	3.8	3.8	3.9	4.2	4.3	4.4
10	5.0	5.0	5.0	5.1	5.2	5.2
∞	5.8	5.8	5.8	5.8	5.8	5.8
B. Four survivors, four non-survivors						
0	0.0	1.0	1.7	2.6	2.9	3.0
2	1.9	2.1	2.5	3.1	3.4	3.5
5	3.8	3.9	4.0	4.3	4.4	4.5
10	5.1	5.1	5.1	5.2	5.2	5.3
∞	5.8	5.8	5.8	5.8	5.8	5.8
C. Survival effect: Panel A minus Panel B						
0	0.0	0.0	0.0	-0.1	-0.1	-0.1
2	0.0	0.0	-0.1	-0.1	-0.1	-0.1
5	0.0	0.0	-0.1	-0.1	-0.1	-0.1
10	0.0	0.0	0.0	-0.1	-0.1	0.0
∞	0.0	0.0	0.0	0.0	0.0	0.0

Table 5

Posterior Mean of Alpha for a Surviving Fund Whose Sample Alpha is 10 Percent, with Residual Cross-Correlations Equal to 0.3

The fund is one of four surviving funds, each of which has a ten-year return history. The other three surviving funds have sample OLS alphas of 1 percent, 0 percent, and -5 percent per annum. All four surviving funds have (i) sample OLS betas of 0.6, 0.3, and 0.1 with respect to the three Fama-French factors MKT, SMB, and HML, (ii) sample residual standard deviations of 17 percent per annum, and (iii) sample correlations of 0.3 between the OLS residuals of any two funds. The true alpha for any fund i (surviving or not) is denoted by α_i and has a prior distribution of the form

$$\alpha_i|\alpha_0 \sim N(\alpha_0, \sigma_\alpha^2), \quad \alpha_0 \sim N(0, \psi^2).$$

The assumed number of non-surviving funds, whose returns are unobserved, is zero in Panel A and four in Panel B. All non-surviving funds are assumed to have a cumulative return for the first five years below -5%, the lowest initial five-year return of the surviving funds in this sample. The values in the table are expressed in percent per year.

σ_α (%)	ψ (%)					
	0	1	2	5	10	∞
A. Four survivors, zero non-survivors						
0	0.0	0.1	0.3	0.9	1.3	1.5
2	1.3	1.4	1.6	2.1	2.5	2.7
5	4.8	4.8	4.9	5.3	5.6	5.8
10	7.8	7.8	7.8	8.0	8.2	8.4
∞	10.0	10.0	10.0	10.0	10.0	10.0
B. Four survivors, four non-survivors						
0	0.0	-0.2	-0.7	-2.3	-3.3	-3.9
2	1.4	1.2	0.6	-0.9	-2.0	-2.4
5	5.1	5.0	4.5	3.1	2.2	1.5
10	8.3	8.3	8.0	7.0	6.1	5.3
∞	10.0	10.0	10.0	10.0	10.0	10.0
C. Survival effect: Panel A minus Panel B						
0	0.0	0.3	1.0	3.2	4.6	5.4
2	-0.1	0.2	1.0	3.1	4.5	5.1
5	-0.4	-0.2	0.4	2.3	3.4	4.4
10	-0.6	-0.5	-0.2	1.0	2.0	3.1
∞	0.0	0.0	0.0	0.0	0.0	0.0

Table 6

Posterior Mean of Alpha for a Surviving Fund Whose Sample Alpha is 10 Percent, with a Prior Link Between Alpha and Residual Risk, and with Residual Cross-Correlations Equal to 0.3

The fund is one of four surviving funds, each of which has a ten-year return history. The other three surviving funds have sample OLS alphas of 1 percent, 0 percent, and -5 percent per annum. All four surviving funds have (i) sample OLS betas of 0.6, 0.3, and 0.1 with respect to the three Fama-French factors MKT, SMB, and HML, (ii) sample residual standard deviations of 17 percent per annum, and (iii) sample correlations of 0.3 between the OLS residuals of any two funds. The true alpha for any fund i (surviving or not) is denoted by α_i and has a prior distribution of the form

$$\alpha_i|\alpha_0 \sim N(\alpha_0, \sigma_\alpha^2), \quad \alpha_0 \sim N(0, \psi^2).$$

When $\gamma > 0$, as in the right-hand columns, the prior for the residual covariance matrix Σ is conditioned on the α (the vector of α_i 's), so that a portion of Σ 's prior mean is proportional to $\alpha\alpha'$. The assumed number of funds with unobserved returns is equal to four throughout, but there is no survival criterion applied in Panel A (i.e., the non-observability is unrelated to realized returns). In Panel B, four non-surviving funds are assumed to have a cumulative return for the first five years below -5%, the lowest initial five-year return of the surviving funds in this sample. The values in the table are expressed in percent per year.

σ_α (%)	$\gamma = 0$			$\gamma = 0.025$		
	$\psi = 2\%$	$\psi = 5\%$	$\psi = 10\%$	$\psi = 2\%$	$\psi = 5\%$	$\psi = 10\%$
A. Four survivors, no survivorship screen						
0	0.4	1.0	1.3	0.4	2.9	2.2
2	1.5	2.1	2.4	2.2	5.8	3.0
5	4.7	5.1	5.4	4.5	9.1	5.8
10	7.7	7.8	8.0	7.6	7.9	8.1
B. Four survivors, four non-survivors						
0	-0.7	-1.8	-2.4	-1.4	-3.3	-3.8
2	0.6	-0.5	-1.1	-0.1	-1.5	-2.4
5	4.2	3.1	2.5	3.6	2.0	1.7
10	7.7	7.0	6.3	7.5	6.7	5.8
C. Survival effect: Panel A minus Panel B						
0	1.1	2.8	3.7	1.7	6.3	5.9
2	0.9	2.6	3.5	2.3	7.3	5.3
5	0.4	2.0	2.9	0.8	7.1	4.2
10	0.0	0.9	1.7	0.1	1.3	2.2

References

- Baks, Klaas P., Andrew Metrick, and Jessica Wachter, 2001, Should investors avoid all actively managed mutual funds? A study in Bayesian performance evaluation, *Journal of Finance* 56, 45–85.
- Brown, Stephen J., and William N. Goetzmann, 1995, Performance persistence, *Journal of Finance* 50, 679–698.
- Brown, Stephen J., William N. Goetzmann, and Roger G. Ibbotson, 1999, Offshore hedge funds: Survival and performance, 1989–95, *Journal of Business* 72, 91–117.
- Brown, Stephen J., William N. Goetzmann, Roger G. Ibbotson, and Stephen A. Ross, 1992, Survivorship bias in performance studies, *Review of Financial Studies* 5, 553–580.
- Brown, Stephen J., William N. Goetzmann, and Stephen A. Ross, 1995, Survival, *Journal of Finance* 50, 853–873.
- Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.
- Carhart, Mark M., Jennifer N. Carpenter, Anthony W. Lynch, and David K. Musto, 2002, Mutual fund survivorship, *Review of Financial Studies* 15, 1439–1463.
- Carlin, Bradley P., and Thomas A. Louis, 1996, *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman and Hall, London.
- Chen, Ming-Hui, Qi-Man Shao, and Joseph G. Ibrahim, 2000, *Monte Carlo Methods in Bayesian Computation*, Springer, New York.
- Chib, Siddhartha, and Edward Greenberg, 1995, Understanding the Metropolis-Hastings algorithm, *American Statistician* 49, 327–335.
- Cochrane, John, 2005, The risk and return of venture capital, *Journal of Financial Economics* 75, 3–52.
- Efron, Bradley, and Carl Morris, 1973, Stein’s estimation rule and its competitors—An empirical Bayes approach, *Journal of the American Statistical Association* 68, 117–130.
- Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.
- Frost, Peter A., and James E. Savarino, 1986, An empirical Bayes approach to efficient portfolio selection, *Journal of Financial and Quantitative Analysis* 21, 293–305.
- Fung, William, and David A. Hsieh, 2000, Performance characteristics of hedge funds and commodity funds: Natural vs. spurious biases, *Journal of Financial and Quantitative Analysis* 35, 291–307.
- Grinblatt, Mark, and Sheridan Titman, 1989, Mutual fund performance: An analysis of quarterly portfolio holdings, *Journal of Business* 62, 393–416.

- Jones, Christopher S., and Jay Shanken, 2005, Mutual fund performance with learning across funds, *Journal of Financial Economics* 78, 507–552.
- Jorion, Philippe, 1986, Bayes-Stein Estimation for Portfolio Analysis, *Journal of Financial and Quantitative Analysis* 21, 279–292.
- Li, Haito, and Yewu Xu, 2002, Survival bias and the equity premium puzzle, *Journal of Finance* 57, 1981–1995.
- Liang, Bing, 2000, Hedge funds: The living and the dead, *Journal of Financial and Quantitative Analysis* 35, 309–326.
- Lunde, Asger, Allan Timmerman, David Blake, 1999, The hazards of mutual fund underperformance: A Cox regression analysis, *Journal of Empirical Finance* 6, 121–152.
- MacKinlay, A. Craig, 1995, Multifactor models do not explain the CAPM, *Journal of Financial Economics* 38, 3–28.
- MacKinlay, A. Craig, and Ľuboš Pástor, 2000, Asset pricing models: Implications for expected returns and portfolio selection, *Review of Financial Studies* 13, 883–916.
- Malkiel, Burton G., 1995, Returns from investing in equity mutual funds 1971 to 1991, *Journal of Finance* 50, 549–572.
- Morris, Carl N., 1983, Parametric empirical Bayes inference: Theory and applications, *Journal of the American Statistical Association* 78, 47–55.
- Pástor, Ľuboš, 2000, Portfolio selection and asset pricing models, *Journal of Finance* 55, 179–223.
- Pástor, Ľuboš, and Robert F. Stambaugh, 1999, Costs of equity capital and model mispricing, *Journal of Finance*, 54, 67–121.
- Pástor, Ľuboš, and Robert F. Stambaugh, 2000, Comparing asset pricing models: an investment perspective, *Journal of Financial Economics*, 56, 335–381.
- Pástor, Ľuboš, and Robert F. Stambaugh, 2002a, Mutual fund performance and seemingly unrelated assets, *Journal of Financial Economics*, .
- Pástor, Ľuboš, and Robert F. Stambaugh, 2002b, Investing in Equity Mutual Funds, *Journal of Financial Economics*, .
- Sharpe, William F., 1994, The Sharpe ratio, *Journal of Portfolio Management* 21, no. 1, 49–58.
- Sims, Christopher A., and Harald Uhlig, 1991. Understanding unit rooters: a helicopter tour. *Econometrica* 59, 1591–1599.