

NBER WORKING PAPER SERIES

BAYESIAN PERFORMANCE EVALUATION

Klaas Baks
Andrew Metrick
Jessica Wachter

Working Paper 7069
<http://www.nber.org/papers/w7069>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 1999

We thank Nick Barberis, Gary Chamberlain, Karsten Hansen, Tom Knox, Tony Lancaster, Andrei Shleifer, and Rob Stambaugh for helpful comments.. The views expressed in this paper are those of the authors and do not reflect those of the National Bureau of Economic Research.

© 1999 by Klaas Baks, Andrew Metrick, and Jessica Wachter. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Bayesian Performance Evaluation
Klaas Baks, Andrew Metrick, and Jessica Wachter
NBER Working Paper No. 7069
April 1999
JEL No. G11, G14, G23

ABSTRACT

This paper proposes a Bayesian method of performance evaluation for investment managers. We begin with a flexible set of prior beliefs that can be elicited without any reference to probability distributions or their parameters. We then combine these prior beliefs with a general multi-factor model and derive an analytical solution for the posterior expectation of “alpha”, the intercept term from the model. This solution can be computed using only a few extra steps beyond maximum likelihood estimation and does not require a comprehensive or bias-free database. We then apply our methodology to a sample of domestic diversified equity mutual funds and ask “what prior beliefs would imply zero investment in active managers?” To justify such a zero-investment strategy, we find that a mean-variance investor would need to believe that less than 1 out of every 100,000 managers has an expected alpha greater than 25 basis points per month. Overall, our analysis suggests that even when the *average* manager is expected to underperform passive benchmarks, it requires very strong prior beliefs to imply zero investment in managers with the *best* past performance.

Klaas Baks
Department of Economics
Brown University
Providence, RI 02912

Andrew Metrick
Department of Economics
Harvard University
Littauer Center, North Yard
Cambridge, MA 02138
and NBER
ametrick@harvard.edu

Jessica Wachter
Department of Economics
Harvard University
Littauer Center, North Yard
Cambridge, MA 02138
wachter@fas.harvard.edu

How should investment managers be evaluated? The standard method of performance evaluation is to compare the returns earned by a manager to a relevant benchmark. In academic practice, this usually involves the regression of manager returns on a set of benchmark returns. To evaluate performance, we then look at the intercept (“alpha”) from this regression and test the null hypothesis that alpha is equal to zero. An alpha greater than zero indicates that a manager adds value relative to the benchmarks. We call this the “frequentist” method of performance evaluation.¹

In this paper, we propose a Bayesian method of performance evaluation. As the name implies, a Bayesian method incorporates prior beliefs about manager performance and combines these beliefs with data. To carry out this analysis, we adopt a flexible family of prior beliefs that is consistent with both behavioral and equilibrium interpretations of factor models. Furthermore, the elicitation of these prior beliefs is simple and intuitive and can be accomplished without any reference to probability distributions or their parameters. The resulting analytical solution for the posterior expectation of alpha can be computed using only a few extra steps beyond maximum likelihood estimation. Using this solution, researchers and decision-makers can see how different prior beliefs would affect inference about alpha. As an application, we analyze a large sample of mutual funds and determine the range of prior beliefs that supports positive investment in non-index managers.

Our approach is similar to four recent papers that use economically motivated prior

¹ Since the seminal work of Jensen (1968), this method has been used in hundreds of papers. For a detailed bibliography of performance evaluation research (most of which employs some form of this method), see <http://www.stern.nyu.edu/~sbrown/performance/bibliography.html>. A final version of this bibliography will be published in a forthcoming special issue of the *Journal of Financial and Quantitative Analysis*.

beliefs and Bayesian updating. Kandel and Stambaugh (1996) show that even statistically insignificant evidence about return predictability can lead rational short-run investors to substantially alter the allocation between stocks and cash in their portfolios, even after taking reasonable prior beliefs into account. Pástor and Stambaugh (1999a) begin with prior beliefs over mispricings in asset-pricing models, and show how firms should then use empirical evidence in their cost-of-equity calculations. Pástor (1999) extends these methods to a portfolio-choice problem over many assets, and uses a flexible set of prior beliefs to reexamine the evidence on the value premium and the home bias. Pástor and Stambaugh (1999b) study portfolio choice under model uncertainty while allowing for mispricings and margin constraints. While our analysis does not take the explicit portfolio-choice perspective that some of these studies do, our motivation and methods are similar to theirs. In particular, our definition of “skill” among managers plays the same mathematical role as “mispricing” does in the latter three papers.²

To demonstrate how our Bayesian approach differs from a frequentist approach in performance evaluation, consider an analogy to the prediction of educational-testing results. Imagine that a student is taking a multiple-choice test. This test has 100 questions and four possible answers for each question. We have access to the prior history of the student’s grades on a series of similar tests. Assume that there is no learning across tests. How should

² There is a related literature that employs Bayesian methods to explore the role of estimation risk on portfolio choice. See Barberis (1999), Bawa, Brown and Klein (1979), Brown (1979), Frost and Savarino (1986), Jobson and Korkie (1980), Jobson, Korkie and Ratti (1979), Jorion (1985, 1986, and 1991) and Klein and Bawa (1976). In these applications, however, prior beliefs about parameters are typically noninformative or come from empirical Bayes procedures. Also, another related line of research focuses on the role of prior beliefs in model testing. See Kandel, McCulloch, and Stambaugh (1995), McCulloch and Rossi (1990), and Shanken (1987).

we estimate the expected score for this student? The standard frequentist approach would be to estimate the average score earned by the student on the previous tests. By definition, this procedure is not influenced by any prior beliefs we may have about the difficulty of the test or the distribution of skill among students.

In a Bayesian framework, we incorporate these prior beliefs. The most obvious illustration of how prior beliefs might matter is for a student whose average score on previous tests was less than 25 percent. We might believe that an expected score below 25 percent is impossible. For example, if the test is not designed to intentionally mislead towards a wrong answer, then there is no reason for a student, on average, to do worse than random guessing over the four choices. In this case, our prior distribution would have no mass below 25 percent, and a student who scored below 25 percent on previous tests would still have an expected score no less than 25.

Another use of prior beliefs would come in the upper tail of the distribution. For example, we may know that this test is so difficult that an expected score above 50 is extremely unlikely. In this case, we have a prior belief for the distribution of actual skill on this test. Then, if we observe a student obtaining an average score above 50 on previous tests, we are likely to adjust this average downwards when forming an expectation for her next test.

In the performance evaluation of investment managers, the traditional techniques are analogous to the frequentist approach of calculating the average score on previous tests. In Bayesian performance evaluation, we acknowledge several similarities between investment management and the testing metaphor just described. First, we recognize that there is likely

to be a class of managers that fall into an “unskilled” group. These managers are analogous to the students who have no useful knowledge for the test. The expected performance of these unskilled managers, before fees and transactions costs, would have a lower bound. At the other extreme, most reasonable prior beliefs would place little weight on an expected return for a manager exceeding its benchmark, by say, 1000 basis points (bp) per year.

Figure 1 shows the shape of our prototypical prior distribution for alpha.³ With probability $1 - q$, a manager is unskilled. Unskilled managers have expected alphas that are negative and with magnitudes approximately equal to the fees that they charge plus the transactions costs that they incur. With probability q , a manager is skilled, and is able to forecast stock returns so as to make profitable trades after costs. This skill translates into the ability to earn alphas in excess of the lower bound, with some managers more skilled than others. Conditional on having skill and taking on a prespecified level of residual risk, prior beliefs on alpha take the shape of the right tail of a normal distribution.

As with any Bayesian analysis, the strength of the results will rely crucially on the plausibility and flexibility of the prior beliefs. Where do these prior beliefs come from? Are they motivated by a reasonable economic model? In Section I, we describe the economic environment in which our analysis takes place. We discuss how the general form of prior beliefs used in our analysis is consistent with both equilibrium and behavioral interpretations of factor models, and we explain the parameters and shape of Figure 1 in more detail.

In Section II, we develop the main theoretical result of the paper. Using prior beliefs

³ Throughout the paper, we follow the convention of Gelman et al. (1995) and use the terms “distribution” and “density” interchangeably.

about alpha with the shape shown in Figure 1, we derive an analytical solution for the posterior expectation of alpha. Our solution is expressed as a formula whose inputs are nothing more than modified moments of well-known distributions. This formula can be applied to managers across a wide variety of contexts: mutual funds, hedge funds, brokerage analysts etc.; one can apply it to a single manager in isolation and it does not require a comprehensive or bias-free database. In each context, the prior beliefs may be different, and rightly so. We also show how prior beliefs can be elicited by intuitive questions such as “what is the probability that a manager has an expected alpha greater than 25 bp per month?”, and we provide the equations that map the answers to these questions into the parameters of the prior belief distribution.

Section III applies our methodology to the performance evaluation of equity-mutual-fund managers. We use a sample of 1437 domestic diversified equity funds in existence at the end of 1996, and look at the full return history for the managers in place at that time. The purpose of this analysis is to show how our methodology can be applied in practice, and how it can be used to better understand manager performance. Using the three-factor model of Fama and French (1993), we calculate the posterior expectation of alpha for each manager over a wide range of prior beliefs, including a baseline case with diffuse prior beliefs for alpha. We then ask, “what prior beliefs would imply zero investment in active managers?” To justify such a zero-investment strategy, we find that a mean-variance investor would have to believe that less than 1 out of every 100,000 managers has an expected alpha greater than 25 bp per month. This result occurs despite a restriction that the average manager has an

expected alpha that is negative and equal in magnitude to expected fees and transactions costs. Section IV summarizes our findings and concludes.

I. Prior beliefs and the economic environment

This section provides a heuristic discussion of the economic intuition underlying Figure 1. Consider an environment where K benchmark portfolios are available, with returns given by F . For equity managers, these portfolios could represent styles such as “large capitalization”, “small capitalization”, “value”, and “growth”. In this environment, suppose that the expected return on any stock can be written as a linear function of these K benchmark (or “factor”) returns and is given by

$$E(r_{it} \mid \Omega_{t-1}) = E(\gamma_{it} + \beta_{it}F_t \mid \Omega_{t-1}). \quad (1)$$

Equation (1) is a standard factor representation for expected returns. r_{it} is the expected excess return on stock i for time t , conditional on information Ω available at time $t - 1$, and β_{it} is a (possibly time-varying) vector of factor loadings. This expected return is a linear function of the expected returns on the benchmark portfolios (F_t) plus an intercept term γ_{it} . We write the intercept term as γ , as opposed to the usual α , so as to avoid confusion with our measure of manager skill. There are many possible interpretations of this equation, but each one shares a common feature with respect to the efficacy of active management. For example, if we interpret equation (1) as the representation of an equilibrium asset-pricing model under which the market is semi-strong efficient (Fama (1970)), then the expectation for γ conditional on (public) information Ω is 0 for all stocks. In this case, a manager who

only has access to information contained in Ω does not have any ability to outperform passive benchmarks. This is an “unskilled” manager in this context. For a manager to outperform the benchmarks, he must have access to information outside the set Ω that enables him to make better forecasts of γ for some stocks. A manager who can make these better estimates of γ is said to have “skill”.

A similar logic applies under different interpretations of equation (1). For example, if the market is not semi-strong efficient with respect to equation (1), then it will be possible to forecast γ using only publicly available information. In this case, we might believe that the probability of skill is higher than it is under semi-strong efficiency, but the definition of skill is still based on the ability to forecast γ . As long as the unconditional expectation of γ is 0 – that is, if equation (1) is correct on average – then there will be some unskilled managers defined simply by their inability to do better than these unconditional estimates. Even if equation (1) does not represent an equilibrium model at all, so that the average premiums earned by the benchmark portfolios are not associated with risk, then we can still think about skilled managers forecasting γ with respect to the benchmarks.⁴ It is these forecasts that will prove crucial for investment decision-making in all of these cases, since managers that can outperform the available set of passive benchmarks will be able to add value to an overall portfolio.

How does the ability to forecast γ translate into abnormal returns earned by the manager? Consider the buy and sell transactions between different managers. When unskilled managers

⁴ For a behavioral interpretation and explanation of the returns to factor loadings, see Daniel and Titman (1997), Lakonishok et al. (1994), and Shleifer (1999).

buy and sell stocks from each other, neither manager has the ability to forecast γ . Thus, the expected γ of the stocks in these transactions is 0. Conversely, skilled managers buy stocks when they forecast that the abnormal returns from trade would (at least) exceed the transactions costs. Thus, when skilled managers buy stocks from unskilled managers, the expected γ of the stocks is positive. Similarly, when an unskilled manager buys stocks from a skilled manager, then the expected γ is negative. Through these actions, skilled managers may reveal some of their information and affect stock prices, but in the process they make profits and cause losses to the unskilled managers.⁵

Using this setup, each manager j would earn an excess return r_{jt} in period t . Now, to evaluate manager performance, we regress these excess returns on the same benchmarks as in (1). This empirical representation is given by

$$r_{jt} = \alpha_j + \beta_j F_t + \varepsilon_{jt}, \quad (2)$$

where α_j is the performance measure for manager j , β_j is a vector of factor loadings, and ε_{jt} is an error term. Typically, α_j is estimated as a fixed parameter that “belongs” to a manager and does not change with his experience or with the size of his portfolio. In principle, these restrictions can be relaxed and α_j can be modelled as a function of these characteristics.⁶

⁵ Given these adverse selection problems, what is the motivation for an unskilled manager to trade? There are many possible explanations: unskilled managers may not know that they are unskilled unless they experiment and make many trades, they may overestimate their ability and believe that they are skilled even after earning losses, or they may be faced with agency problems and career concerns that encourage trading. We assume that skilled managers face similar issues, so that skilled managers, while earning positive abnormal returns on some trades, still may earn negative abnormal returns on others.

⁶ It would also be possible to estimate a conditional factor model and let the β_j vector vary over time, but this would require a bit more structure than we present here. If managers can successfully forecast factor realizations (F_t) and alter their factor loadings in response, then such a structure would be necessary. See Ferson and Schadt (1996) for an application of conditional factor models to performance evaluation.

For the remainder of this discussion, we drop the j and t subscripts.

In frequentist performance evaluation, the next step is to assess the statistical significance of the $\hat{\alpha}$ estimate under the null hypothesis that α is 0. This is the point of departure for Bayesian performance evaluation. For now, consider the case where the distribution of ε_t is known, so that the manager's level of residual risk is held constant.⁷ Then, given the economic framework sketched above, we begin with a prior distribution for α with a shape given by Figure 1, with a prior probability q that a manager is skilled and probability $1 - q$ that he is unskilled. The key features of this distribution are the point mass at $\underline{\alpha} < 0$, and the right tail of a normal distribution as the functional form in the skilled state. The normal distribution is chosen for analytic tractability, but is also a useful approximation for skill distributions in many other contexts. The point mass occurs at $\underline{\alpha} = a - fee - cost < 0$. Thus, an unskilled manager is expected to earn a negative alpha that consists of three components: a , his losses due to transactions with skilled investors, fee , his total fees, and $cost$, his transactions costs. The first component, a , is the same for any manager. The other components, fee and $cost$, will generally differ across managers.⁸

Why do we assume a lower bound at $\underline{\alpha}$? Under some interpretations of equation (1), this assumption is logical. For example, if the market is semi-strong efficient with respect to (1), then no manager should be expected to have an alpha below $\underline{\alpha}$. While many managers will

⁷ In Section II, we introduce a link between manager's residual risk and prior expectations for α .

⁸ Here, the fee is a fixed percentage that is independent of the returns earned by the manager. For some applications, such as mutual funds, this assumption is reasonable. For hedge funds, however, fees typically include a fraction of the portfolio return. In practice, this complication is not a major problem. When we do an empirical implementation, we can elicit priors for α before fees (as in Section III); then, more complicated fee schedules can be brought in after the calculation of posterior beliefs.

have *realizations* below this level, one would need to be systematically acquiring non-public “misinformation” in order to have an *expected* α below $\underline{\alpha}$. If, instead, the market is not semi-strong efficient and managerial skill is based upon the exploitation of decision-making biases using public information, then differential incidence of such biases would result in a some prior mass below $\underline{\alpha}$.⁹ By imposing a lower bound at $\underline{\alpha}$, we are assuming that any behavioral biases are evenly distributed among all managers. While this assumption will affect inference for the worst-performing managers, it should have little effect on posterior beliefs for the best-performing ones. This point is discussed at the end of Section III.

II. Bayesian performance evaluation

This section, along with the Appendix, provides the details of our methodology. In Section II.A, we provide the likelihood function for a general (unconditional) factor representation of manager returns. Section II.B gives a complete mathematical representation for prior beliefs, and Section II.C poses three questions necessary to elicit this representation. In Section II.D, we combine these prior beliefs with the likelihood function and derive an analytical solution for the posterior expectation of alpha. Section II.E discusses how our solution can be used for portfolio choice.

A. Likelihood

Let r denote a $T \times 1$ vector of excess returns for a manager and F a $T \times K$ matrix

⁹ One way to model this possibility would be to make the prior distribution for alpha symmetric around $\underline{\alpha}$. In fact, this assumption would greatly simplify our analysis. We do not use a symmetric distribution because we find such prior beliefs to be implausible. A symmetric prior distribution for α would imply that for every skilled manager with superior judgement or the ability to exploit the behavioral anomalies of other investors, there is another manager who systematically does the opposite.

of factor returns. The regression disturbance ε_t in equation (2) is assumed to be a serially uncorrelated, homoskedastic realization from a normal distribution, with zero mean and variance equal to σ^2 . Then, we write the likelihood for r conditional on F as

$$p(r \mid \alpha, \beta, \sigma^2, F) = N(\alpha \iota_T + F\beta, \sigma^2 I_T), \quad (3)$$

where ι_T is a T -vector of ones, and I_T is a $T \times T$ identity matrix. Thus, manager returns conditional on factor returns are normally distributed and have a standard factor structure. We assume that the factors F do not depend on α , β or σ , so the exact specification of the factor likelihood is not necessary for our analysis in this section.

B. Prior Beliefs

The next step is to state the prior beliefs for the parameters in (3). As discussed in the previous section, managers are either skilled or unskilled. These two states of the world are indexed by the state variable Z , with $Z = 1$ denoting the skilled state and $Z = 0$ denoting the unskilled state. The probability of the skilled state is q . In the absence of data on the returns of a manager, the probability that the manager has skill is simply given by the probability that skill exists in the population, or $P(Z = 1)$. We assume that only α depends on whether or not the manager is skilled; the factor loadings, β , and residual risk, σ , do not.

Under these assumptions, the prior distribution can be written as

$$p(\alpha, \beta, \sigma^2) = [p(\alpha \mid Z = 0)P(Z = 0) + p(\alpha \mid Z = 1)p(Z = 1)]p(\beta, \sigma^2). \quad (4)$$

We use a diffuse prior on β and σ^2 (Gelman et al. (1995)):

$$p(\beta, \sigma^2) \propto \frac{1}{\sigma^2}. \quad (5)$$

The diffuse prior on β is necessary in order to obtain analytical results. It is also a reasonable starting point when analyzing managed portfolios, where β can be estimated relatively precisely (as compared to the β of individual stocks). The diffuse prior on σ^2 is not necessary for analytical results, but it simplifies notation and allows us to focus our attention on the role played by α . Appendix A relaxes this second assumption and solves for the posterior when σ^2 has an informative prior; then, the diffuse prior used here becomes a limiting case.

We turn next to α , the main parameter of interest.¹⁰ Essentially, we want to write down a mathematical representation of Figure 1. The one additional twist is to recognize that prior beliefs for α should be conditioned on some level of residual risk. For example, consider a fully-invested manager who has an α of α_j and is taking on s units of residual risk. Then, if this manager were to take on a fifty-percent cash position, his residual risk would decrease to $s/2$ and his α would fall to $\alpha_j/2$. The full specification of the prior for α recognizes this relationship and is written as

$$P(Z = 1) = q, \quad (6)$$

$$P(Z = 0) = 1 - q, \quad (7)$$

$$p(\alpha | Z = 0, \sigma^2) = \delta_{\underline{\alpha}}, \quad (8)$$

$$p(\alpha | Z = 1, \sigma^2) = 2N\left(\underline{\alpha}, \sigma_{\alpha}^2 \left[\frac{\sigma^2}{s^2}\right]\right) \mathbf{1}_{\alpha > \underline{\alpha}}. \quad (9)$$

¹⁰ In addition to the informed prior on α that is given below, Section III also considers a baseline case where α has a diffuse prior. The posterior distribution for this case is solved in Appendix B.

where δ_x is the Dirac delta function with mass point at x , $\mathbf{1}_X$ is the indicator function for the set X , $\underline{\alpha}$ is a negative constant representing the expected α for an unskilled manager, and s^2 is an arbitrary constant specified by the researcher before priors are elicited. Finally, we assume that the parameters of the (unspecified) factor likelihood are independent of α , β , and σ .

The ratio $\frac{\sigma^2}{s^2}$ effectively links the posterior distributions of σ and α . As discussed above, this link allows us to adjust for the fact that a skilled manager can control his expected α through the strategic use of leverage.¹¹ The importance of this relationship becomes more clear in the next section when we discuss the elicitation of priors.

The prior link between α and σ is first suggested by MacKinlay (1995) and is implemented in Pástor (1999), Pástor and Stambaugh (1999a), and Pástor and Stambaugh (1999b). Mathematically, our link is identical to theirs, although their motivation is somewhat different. In these papers, σ_α is an index of potential “mispricing”, and the motivation for the link is to reduce the ex-ante probability of very high Sharpe ratios among portfolios that combine benchmark and non-benchmark assets.

C. Elicitation of Prior Beliefs

The parameters of the prior belief distribution may not be natural or intuitive objects for many researchers/investors. It is possible, however, to elicit prior beliefs using questions that make no reference to the parameters of probability distributions. Instead, the questions focus on probabilities that a manager has an expected (or “true”) α that exceeds specific

¹¹ The same argument can be used to motivate a link between β and σ . Since we use a diffuse prior for β , such a link is not applicable.

levels. For example, consider any specific factor representation for equation (2); the choice of factors is left to the researcher. Then, given this factor representation, assume that the manager under study has a residual variance, σ^2 , equal to a specific value. Again, the exact level of this value is not important, and can be chosen by the researcher. We call this level s^2 , and it serves as the constant denominator term in equation (9). Then, conditional on $\sigma^2 = s^2$ and the chosen factor representation, the researcher should answer the following three questions:

1. What is the probability that the manager has an α greater than 10 bp? (Call this answer $q(10)$)
2. What is the probability that the manager has an α greater than 25 bp? (Call this answer $q(25)$)
3. What is the α for the manager if he is “unskilled”? (Call this answer $\underline{\alpha}$)

In answering these questions, the researcher should not consider any return-based information about the manager that coincides with the sample period under study; this forces the exclusion of all information about the length of time the manager has survived or the level of assets that he has under management, both of which will tend to be correlated with past returns. Instead, the answers should reflect a thought experiment about a new manager before any return information has been observed. It is acceptable to consider information about the manager that precedes any of the returns in the sample period. For example, we may know that a hedge-fund manager needs to raise significant funds – perhaps by pointing to a previous record – before his current return series begins. This information could lead

the researcher to higher prior probability of skill for hedge fund managers than for managers in sectors with different entry standards.

Note that questions (1) and (2) are not asking about the expected realization of $\hat{\alpha}$, but about “true” values of α . Realizations of $\hat{\alpha}$ will depend on sampling variability. With enough sampling variability, the fraction of managers with an $\hat{\alpha}$ above 25 bp could be very close to 0.5, even if no managers are skilled. True α , on the other hand, is the $\hat{\alpha}$ we would expect as the number of time periods goes to infinity. Thus, if a researcher believes that no managers are skilled, then $q = 0$, and thus $q(10)$ and $q(25)$ should be 0 as well.

The answer to question 3 must take into account expected fees and costs for the manager. In addition, the researcher must include expected losses due to transactions with skilled managers, the parameter that we call a in Figure 1. As long as the total value of portfolios managed by skilled managers is “small” relative to the market, there is little reason to believe that a would be very different from zero. Below, we provide a formula for calculating a baseline value for a as a function of $q(10)$ and $q(25)$.

All of three of these questions are conditioned on a specific level of residual variance, s^2 . This conditioning is crucial, and we believe that prior beliefs on α are not well-defined without it. The same reasoning discussed in Section II.B also applies here: if a manager has an expected α of α_j when his residual standard deviation is s , then his expected α would be $\alpha_j/2$ if he levered down his portfolio and took on only $s/2$ units of residual risk. By including the σ^2/s^2 term in the prior beliefs for α , we link our beliefs for α and σ in a way consistent with our elicitation procedures. In this respect, the prior beliefs elicited through

these questions are not really about α , but are instead about Sharpe ratios for combinations of the manager and the benchmarks.¹² In this way, we return to the original motivation for the link as given by MacKinlay (1995). To be completely consistent, we should also link our prior beliefs for a (and, by extension, $\underline{\alpha}$) to the ratio σ^2/s^2 . Unfortunately, this link is not tractable. As long as a is small, however, the omission will not be quantitatively important.

Given the answers to these questions, we have $\underline{\alpha}$ (from question 3). Let $\Phi(x)$ denote the cdf of a standard normal distribution evaluated at x . Then, we have two equations,

$$q(25) = P(\alpha > 25 \mid \sigma^2 = s^2) = 2q \left(1 - \Phi \left(\frac{25 - \underline{\alpha}}{\sigma_\alpha} \right) \right), \quad (10)$$

and

$$q(10) = P(\alpha > 10 \mid \sigma^2 = s^2) = 2q \left(1 - \Phi \left(\frac{10 - \underline{\alpha}}{\sigma_\alpha} \right) \right), \quad (11)$$

which we can solve for the two unknowns, q and σ_α . Note that our use of 10 and 25 bp is arbitrary, and any other two points would also suffice.

In some applications (such as Section III), it is useful to set a baseline case where the expectation of α , conditional on $\sigma = s$, is equal to 0. This implies a closed system where all managers' profits must cause offsetting losses for other managers. In this case, we change question 3 to ask only about expected fees and costs ($fee + cost$), and we then include a third equation

¹² We are grateful to Rob Stambaugh for suggesting this interpretation.

$$\underline{\alpha} = -q\sigma_{\alpha}\sqrt{\frac{2}{\pi}} - fee - cost, \quad (12)$$

where the first term on the right-hand-side of (12) represents the value of a such that the expectation of α equals $-fee - cost$ conditional on $\sigma = s$. We can then solve (10) - (12) for the three unknowns: q , σ_{α} , and $\underline{\alpha}$. Table I illustrates some solutions to this system: given inputs of $q(25)$, $q(10)$, fee , and $cost$, we provide the solutions for q , σ_{α} , a , and $\underline{\alpha}$. In most of the examples in the table, a tends to be very small, so $\underline{\alpha}$ is close to $-fee - cost$. Alternatively, one can elicit $q(10)$ and $q(25)$ before fees. To compute the prior parameters under this alternative method, we just replace $\underline{\alpha}$ with $a - cost$ on the right-hand-side of (10) and (11). This is the elicitation method used in Section III.

We impose the requirement that the answers to the elicitation questions be real values and not distributions. Thus, prior beliefs for the parameters $\underline{\alpha}$, σ_{α} , and q , will be real values as well. This requirement allows us to obtain an analytical solution for the posterior mean of α . If, instead, the answers are given as distributions, then we will need to solve numerically for the posterior mean of α .

D. Posterior Beliefs

Our goal in this section is to calculate the mean of the posterior distribution for α . Most of the intuition for this solution is contained in Figure 2 and its corresponding notation as developed in (13) - (21). The details are given beginning with (22) and in Appendix A.

We denote the mean of the posterior distribution for α , $E[\alpha | r, F]$, as $\tilde{\alpha}$. Similarly, we denote the posterior probability that a manager is skilled, $P(Z = 1 | r, F)$, as \tilde{q} . Then, it

follows that we can write $\tilde{\alpha}$ as

$$\tilde{\alpha} = \tilde{q}E[\alpha | Z = 1, r, F] + (1 - \tilde{q})\underline{\alpha}. \quad (13)$$

The first term on the right-hand-side of (13) reflects the contribution to the posterior mean coming from the possibility that the manager has skill; the posterior probability of skill (\tilde{q}) is multiplied by the posterior expectation of α conditional on skill. The second term on the right-hand-side of (13) reflects the contribution coming from the possibility that manager is unskilled; the posterior probability ($1 - \tilde{q}$) is multiplied by $\underline{\alpha}$. Note that our procedure is not updating q for the “population”; q is treated as *known*. Rather, our inference problem only concerns the specific manager under study.

To calculate $\tilde{\alpha}$, we need to solve for the two unknown elements on the right-hand-side of (13): $E[\alpha | Z = 1, r, F]$ and \tilde{q} . The problem of computing $\tilde{\alpha}$ is thus considered in two parts. First, we calculate $E[\alpha | Z = 1, r, F]$, the expectation conditional on skill. Except for the truncation at $\underline{\alpha}$, this problem is very similar to that solved in Pástor and Stambaugh (1999a). In fact, it is easier because of the diffuse prior on β , and closed-form solutions for the posterior distribution can be obtained. Second, we calculate \tilde{q} , the posterior probability that the manager is skilled. Both pieces can be solved for in terms of maximum likelihood estimates and the pdf and cdf of the t -distribution.

In expositing our solution, it is helpful to introduce some notation:

$$X \equiv \begin{pmatrix} \iota_T & F \end{pmatrix}, \quad (14)$$

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \equiv (X'X)^{-1}X'r, \quad (15)$$

$$m \equiv \text{top left element of } (X'X)^{-1}, \quad (16)$$

$$\text{var}(\hat{\alpha}) \equiv m\sigma^2. \quad (17)$$

Then, the posterior of α conditional on the variance σ and $Z = 1$, which we call the “skilled posterior distribution”, is given by a truncated normal distribution:¹³

$$p(\alpha \mid Z = 1, r, F, \sigma^2) \propto N(\alpha', \sigma'^2) \mathbf{1}_{\alpha > \underline{\alpha}} \quad (18)$$

where

$$\alpha' = \lambda \hat{\alpha} + (1 - \lambda) \underline{\alpha}, \quad (19)$$

$$\sigma'^2 = \left(\frac{1}{\text{var}(\hat{\alpha})} + \frac{1}{\sigma_\alpha^2 \left(\frac{\sigma^2}{s^2} \right)} \right)^{-1}, \quad (20)$$

$$\lambda = \frac{\sigma'^2}{\text{var}(\hat{\alpha})}. \quad (21)$$

Equations (18) - (21) are illustrated graphically in Figure 2. α' is the mode of the skilled posterior distribution given in (18); it would also be the mean, and σ'^2 the variance, of the untruncated version of this distribution. α' is written in (19) as a weighted average of the maximum likelihood estimate ($\hat{\alpha}$) and the prior mode ($\underline{\alpha}$), with weights given by λ and

¹³ See Appendix A for the details of these calculations.

$1 - \lambda$, respectively. In (20), $\text{var}(\hat{\alpha})$ represents the variance (in a frequentist sense) of the maximum likelihood estimate for α , conditional on a known residual variance of σ^2 . The posterior precision, $1/\sigma'^2$, is the sum of the precision of the prior and the precision of the data. Intuitively, this says that after having observed the data, there is greater certainty about the location of the posterior distribution of α than there was for the prior. Thus, the weight λ is determined in (21) by the relative precision of prior beliefs versus sample information. The greater the precision of $\hat{\alpha}$, the more the mode is shifted towards $\hat{\alpha}$ and away from the prior mode $\underline{\alpha}$.

The marginal posterior for α (conditional on skill) can be obtained in closed form by successively integrating out β and σ from the joint posterior. Once β is integrated out, the model resembles one where normal data is combined with a conjugate prior. Therefore, familiar techniques (see, e.g. Gelman et al. (1995)) can be used to integrate out σ , suitably adjusted to reflect the truncation at $\underline{\alpha}$. The marginal distribution is then given by:

$$p(\alpha | Z = 1, r, F) \propto t_\nu \left(\alpha', \frac{\lambda m h}{\nu} \right) \mathbf{1}_{\alpha > \underline{\alpha}}, \quad (22)$$

where

$$h = S + \frac{1 - \lambda}{m} (\hat{\alpha} - \underline{\alpha})^2, \quad (23)$$

$$\hat{\theta} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}, \quad (24)$$

$$S \equiv (r - X\hat{\theta})' (r - X\hat{\theta}), \quad (25)$$

$$\nu = T - K. \quad (26)$$

Much of the intuition of Figure 2 carries through here. Roughly speaking, s^2/σ_α^2 is a measure of prior precision and $1/m$ is a measure of precision of the data (recall that m is the upper-left corner of $(X'X)^{-1}$). The posterior mode α' , is once again a weighted average of the prior mode and the maximum-likelihood estimate. λ has the same interpretation here as in (21); the greater the precision of the data relative to the prior precision, the more the posterior mode is shifted towards the maximum-likelihood estimate.

The posterior expectation of α in the skilled state can then be calculated as

$$E(\alpha \mid Z = 1, r, F) = \alpha' + \frac{\lambda m h}{\nu - 2} t_{\nu-2} \left(\underline{\alpha}; \alpha', \frac{\lambda m h}{\nu - 2} \right) \frac{1}{\int_{\underline{\alpha}}^{\infty} t_{\nu} \left(\alpha; \alpha', \frac{\lambda m h}{\nu - 2} \right) d\alpha}. \quad (27)$$

The first term on the right-hand-side of (27) is just the mode of the skilled posterior distribution (the mean of the t -distribution), while the second term is an adjustment for the truncation at $\underline{\alpha}$. We use the notational convention that $t_{\nu}(\alpha; x, y)$ is a t -distribution evaluated at α with mean x , variance y and ν degrees of freedom.

We next solve for the second unknown element in (13), \tilde{q} , the posterior probability that the manager is skilled. From Bayes' formula for binomial variables, it follows that

$$\tilde{q} \equiv P(Z = 1 \mid r, F) = \frac{p(r \mid Z = 1, F) q}{p(r \mid Z = 1, F) q + p(r \mid Z = 0, F) (1 - q)}. \quad (28)$$

Dividing through by $p(r \mid Z = 1, F)$ yields

$$\tilde{q} = \frac{q}{q + \frac{1-q}{B}}, \quad (29)$$

where B is a likelihood ratio given by

$$B \equiv \frac{p(r \mid Z = 1, F)}{p(r \mid Z = 0, F)}. \quad (30)$$

If observing the realized data is equally likely whether the manager is skilled or unskilled, then $B = 1$, and our posterior probability of $Z = 1$ equals the prior probability: $\tilde{q} = q$. The more likely the data are for a skilled manager relative to an unskilled manager, the higher is B and thus, the higher is \tilde{q} .¹⁴

As shown in Appendix A,

$$B = \frac{t_{\nu-1}(\underline{\alpha}; \hat{\alpha}, \frac{S}{\nu-1}(m + \frac{\sigma_{\alpha}^2}{s^2}))}{t_{\nu-1}(\underline{\alpha}; \hat{\alpha}, \frac{S}{\nu-1}m)} \left(2 \int_{\underline{\alpha}}^{\infty} t_{\nu} \left(\alpha; \alpha', \frac{\lambda m h}{\nu} \right) d\alpha \right). \quad (31)$$

This likelihood ratio is a ratio of two t -distributions, multiplied by a term to correct for the truncation at $\underline{\alpha}$.

Finally, with values for $E(\alpha | Z = 1, r, F)$ (equation (27)) and \tilde{q} (equations (29,31)), we can substitute into (13) and obtain a solution for $\tilde{\alpha}$.

E. Portfolio Choice

One application for our approach is to a portfolio-choice problem for an investor choosing among a single manager and a set of benchmark portfolios. This analysis requires a predictive joint distribution for r and F . Appendix C derives the remaining components of this distribution and describes how to sample from it. A complete portfolio-choice analysis also requires a specific utility function and a set of trading institutions (e.g., margin requirements, transactions costs) and is outside the scope of our paper.¹⁵ There are, however,

¹⁴ Note that by dividing the numerator and denominator of (29) by q , the the posterior odds ratio $\frac{1-q}{q} \frac{p(r|Z=0,F)}{p(r|Z=1,F)}$ for testing the hypothesis $H_0 : q = 0$ versus $H_1 : q > 0$ appears. Thus, B is a ‘‘Bayes factor’’ associated with testing H_0 versus H_1 .

¹⁵ For examples of portfolio-choice analysis in environments similar to the one studied here, see Pástor (1999) and Pástor and Stambaugh (1999b). For general discussion of portfolio-choice decision rules in a Bayesian framework, see Bawa, Brown and Klein (1979) and Chamberlain (1998).

simple portfolio-choice questions that can be answered using only the posterior mean, $\tilde{\alpha}$. For example, consider a mean-variance investor choosing among the benchmark portfolios and a single manager. Assume that this investor is facing a single-period problem, faces no taxes or transactions costs, and can sell short the benchmark portfolios with no margin restrictions. Then, under what condition would the investor choose positive investment in the manager? In an environment where α is known, this condition is the simple $\alpha > 0$ (Gibbons et al. (1989)). When α is unknown, as in our problem, the condition can often be reduced to the analogous $\tilde{\alpha} > 0$; that is, a positive posterior expectation for α . (Pástor (1999). Appendix D shows that this condition holds in our framework as well. We make use of this condition in next section.

III. Application: Actively-managed mutual funds

By the late 1990s, there was a wide variety of passive benchmark portfolios available to U.S. investors in the form of index funds, and such funds typically have lower fees and transactions costs than their active counterparts. While considerable evidence exists that the majority of managers underperform these passive benchmarks (after fees), it is difficult to say anything conclusive about expected performance of the “best” managers. When posed as a frequentist test of the null hypothesis that $q(25)$ is 0 against the alternative that $q(25)$ is 0.001, the current models, methods, and data are insufficient to give the tests much power. In practice, investors act as if $q(25)$ is positive, the highest-performing managers attract large inflows of new funds, and there is an ongoing academic debate about whether such

behavior is justified.¹⁶

In this section, we use our Bayesian methodology to take an investor’s perspective and ask, “given the evidence, what range of prior beliefs would imply positive investment in any active mutual-fund manager?” To answer this question, we use the $\tilde{\alpha} > 0$ condition discussed in Section II.E. Despite several conservative assumptions, the positive investment range includes many points that, in a frequentist sense, are nearly indistinguishable from $q(25) = 0$. We do not claim that this analysis could be used to make actual portfolio choices in a world of taxes, loads, and margin restrictions. Rather, it is an attempt to demonstrate our methodology while viewing the mutual-fund evidence through a new lens.

Our data is drawn from the Center for Research in Security Prices (CRSP) mutual-fund database (CRSP, 1999). This database includes information collected from several sources and is designed to be a comprehensive sample of all mutual funds from 1963 to 1996. We restrict ourselves to the subset of funds still operating at the end of 1996, and only include returns that have been earned by current (as of December 1996) managers. We include only the returns earned by current managers because we interpret α as a fixed parameter that is a characteristic of managers, not of funds. In the remainder of the section, we use the terms “fund” and “manager” interchangeably. We include team-managed funds only if a name is provided for at least one member of the team; returns for such funds are included for the tenure of the team’s longest-standing member. Furthermore, we restrict ourselves to funds

¹⁶ For some recent papers in this debate, see Brown and Goetzmann (1995), Carhart (1997), Chevalier and Ellison (1999), Daniel et al. (1997), Gruber (1996), Hendricks et al. (1993), Lakonishok et al.(1992), Malkiel (1995), and Zheng (1999).

with at least one complete year of return history. The resulting sample includes 1437 funds with an average of 51 months of return history. This sample suffers from survivor bias and will not be representative of mutual-fund performance as a whole; this bias is not a problem for our analysis.

The next step is to choose a set of benchmarks for the evaluation. Since there is no consensus on the “correct” asset-pricing model, most performance-evaluation studies use several different models in their analysis. Since our purpose is more to demonstrate our methodology than to provide a definitive analysis of mutual funds, we restrict ourselves to a single well-known model – the three-factor model of Fama and French (1993). The model is given by

$$r_{jt} = \alpha_j + \beta_1 RMRF_t + \beta_2 SMB_t + \beta_3 HML_t + \varepsilon_{jt}, \quad (32)$$

where r_{jt} is the excess return to fund j in year t , α_j is the performance measure, and $RMRF_t$, SMB_t , and HML_t are the time t returns to benchmark portfolios constructed using market, size, and value strategies.¹⁷ While there is an ongoing debate about whether the coefficients in this regression are proxies for priced risk, we take no position on this issue and simply view the three-factor model as a method of performance attribution. Thus, we interpret the estimated alphas as abnormal returns in excess of what could have been achieved by a matched investment in the benchmark portfolios. This model suits our purposes, as we wish to determine whether a manager can outperform an available set of passive index funds,

¹⁷ See Fama and French (1993) for details on the construction of these portfolios. We are grateful to Ken French for providing the factor returns.

and the main style categories for both indexation and active management are along size and value/growth dimensions; while the benchmark returns in (32) are not themselves available as passive index funds, they are very similar to (combinations of) index products available in the late 1990s.¹⁸ The fact that we ignore any transactions costs that would be incurred in constructing these benchmark portfolios is keeping with the conservative bias of our analysis; inclusion of such transaction costs would make managers look better.

Rather than elicit specific prior beliefs, we construct a mapping from a wide range of prior beliefs to posterior beliefs. To simplify the analysis, we elicit priors before fees so that we can use the same interpretation of q and σ_α for all managers. In principle, we could use different prior parameters for each manager depending on their style, education, or other characteristics.¹⁹ Total fees are reported in the database and vary across managers and across time. Consistent with our elicitation, we analyze gross returns (by adding back fees for each year), and then subtract the current *fee* at the end. Transactions costs are not reported; we use a single value, 6 bp per month, as the *cost* for every manager.²⁰ For the denominator of the leverage term, given as s^2 in equation (9), we use the cross-sectional mean of $\hat{\sigma}^2$ from our sample ($= 0.00029$). This level of s^2 is a useful normalization that

¹⁸ Low-cost index funds are available in 1999 from the Vanguard mutual-fund family (among other places) in large-capitalization value, small-capitalization value, large-capitalization growth, and small-capitalization growth categories. Note that low-cost “momentum” index funds are not available, since momentum investing is by nature a high-turnover activity. This is the main reason we do not include a momentum benchmark in our analysis.

¹⁹ Chevalier and Ellison (1999) provide evidence that many such characteristics are correlated with alphas.

²⁰ This value roughly corresponds to the average monthly transactions costs for mutual funds and large institutions found in other studies; see Carhart (1997) for turnover rates and implied trading costs, Keim and Madhavan (1997) for per-trade costs, and Perold (1988) for the methodology behind these calculations. We discuss alternative assumptions at the end of this section.

makes it easier to interpret the results, since prior beliefs can then be stated relative to an average level of residual risk. Finally, to set $\underline{\alpha}$, we impose a conservative assumption that, conditional on $\sigma^2 = s^2$, the expectation of α is $-fee - cost$. This value of $\underline{\alpha}$, which will vary across funds, is calculated as in equation (12) from Section II.C. At the end of this section, we discuss the implications of changing some of these assumptions. Note that we cannot use our data to infer anything about the “true” values of q, σ_α , or $\underline{\alpha}$; this kind of analysis would require an unbiased sample (ours is not) and an additional adjustment for dependencies across managers. Thus, we are not attempting to say what prior beliefs “should be”. Rather, our goal is to better understand how these prior beliefs matter for inference and decision-making.

Given prior beliefs, the next step is to combine these beliefs with the data and calculate posterior beliefs. In addition to the informed prior beliefs for α (as in Section II.B), we also consider a baseline case with diffuse prior beliefs for α (as in Appendix B). As an example, consider the Oakmark fund. From a frequentist perspective, this is one of the best performing funds in our sample. Since the present manager’s (Robert Sanborn) returns began in September 1991, the fund has earned an $\hat{\alpha}$ of 102 bp per month (before fees). The standard error on this alpha estimate is 24 bp, resulting in a t -statistic of 4.27. If we perform a frequentist test of the null hypothesis that α is zero, we obtain a p -value of 0.00003; thus, such a performance would be obtained by chance about 3 times for every 100,000 managers. Even after considering Oakmark’s 1996 total fees of 9.8 bp per month, we could still reject the null hypothesis that $\alpha = 0$ (after fees) at a very high level of confidence. Not surprisingly,

the baseline case using a diffuse prior for α yields an $\tilde{\alpha}$ of 92.2 bp per month; a posterior mean that is identical to the maximum likelihood estimate (before fees) minus the total fees.

How different are posterior beliefs when using an informed prior for α ? Informed prior beliefs tend to shrink $\tilde{\alpha}$ towards its prior mean. Combining the fees of 9.8 bp with our assumption that transactions costs are six bp, this prior mean for Oakmark's expected α is equal to -15.8 bp.²¹ Figure 3 plots level curves for $\tilde{\alpha}$, the posterior expectation of α , as a function of q and σ_α . The figure includes level curves at $-10, 0, 10, 20, 30, 40,$ and 50 bp. To ease the interpretation of this figure, we can translate points from (q, σ_α) space to implied probabilities for $q(25)$. Here, the proper interpretation of $q(25)$ is the probability of α greater than 25 basis points per month, after transactions costs but before fees, and conditional on $\sigma^2 = 0.00029$. For example, the point $(0.01, 19)$ lies on the level curve for $\tilde{\alpha} = 0$. At this point, the implied $q(25)$ can be calculated as 0.0014. Similarly, the point $(0.1, 13)$ on the same $\tilde{\alpha} = 0$ frontier has an implied $q(25)$ of 0.0013.

Instead of mapping prior beliefs for q and σ_α into $\tilde{\alpha}$, we could also take a more intuitive approach and go directly from $q(25)$ to $\tilde{\alpha}$. To do so, we imagine that the investor makes the following request: "I know that $q(25)$ is equal to x , but I have no opinion about $q(10)$ or any other point in the distribution. Please tell me the *minimum* possible value of $\tilde{\alpha}$ that is consistent with $q(25)$ equal to x ." To carry out this analysis, we first compute the values of $q, \sigma_\alpha,$ and $\underline{\alpha}$ that are consistent with any given level of $q(25)$ and with our assumption that the prior expectation of α is $-fee - cost$ when $\sigma^2 = s^2$. Then, for every level of $q(25)$, we

²¹ All statements about prior expectations are made conditional on $\sigma^2 = 0.00029$.

calculate the minimum value of $\tilde{\alpha}$ over all possible combinations of q , σ_α , and $\underline{\alpha}$; we denote these minimum values as $\tilde{\alpha}_{\min}(q(25))$.

Figure 4 plots $\tilde{\alpha}_{\min}(q(25))$ for the Oakmark fund. For extremely small values of $q(25)$, even very high abnormal returns will not lead the investor to update his beliefs very much, and $\tilde{\alpha}_{\min}$ will be negative and close to $\underline{\alpha}$. $\tilde{\alpha}_{\min}$ first becomes positive at $q(25) = 0.0014$; this is close to the point (0.01,19) in Figure 3. Thus, if the investor believes that there is a prior probability of 0.14 percent that a manager will have an α , before fees, that is at least 25 bp per month, then any combination of q and σ_α consistent with this probability would imply a positive $\tilde{\alpha}$ for Oakmark. Above this point, $\tilde{\alpha}_{\min}$ continues to increase with $q(25)$, and stands at over 80 bp when $q(25)$ is 0.05. This is not surprising, since the posterior belief that Oakmark has skill (\tilde{q}) will be very close to 1 for such high values of $q(25)$; in this case, $\tilde{\alpha}$ is not that far from its frequentist estimate $\hat{\alpha}$.

Figures 5 and 6 extend this analysis to all of the managers in our sample. The figure plots the highest and tenth highest $\tilde{\alpha}_{\min}$ among all managers for every level of $q(25)$: Figure 5 plots the range from $q(25) = 0$ to $q(25) = 0.001$; Figure 6 plots the range from $q(25) = 0.001$ to $q(25) = 0.01$. As a comparison, if we use a diffuse prior for α , approximately half of the managers have positive values of $\tilde{\alpha}$, and 20 managers have an $\tilde{\alpha}$ greater than three times the size of $\hat{\sigma}$ – the equivalent of a frequentist t -statistic above 3. As seen in Figure 5, the best performing manager's $\tilde{\alpha}_{\min}$ becomes positive at about $q(25) = 0.00001$, and the tenth best manager becomes positive at $q(25) = 0.00056$. At $q(25) = 0.01$, the best performing manager has an $\tilde{\alpha}_{\min}$ of about 30 bp, and the tenth best performing manager has an $\tilde{\alpha}_{\min}$ of

about 19 bp. Note that Oakmark is never among the top ten over this range. In general, the best performing managers at low levels of $q(25)$ are those with a positive and “significant” $\hat{\alpha}$ and a long history of returns. In these cases, the posterior distributions for α , β , and σ are relatively tight compared to those for managers with shorter histories. Other things equal, tight posterior distributions for all of the parameters will tend to increase $\tilde{\alpha}$ for good performers.

What have we learned from this exercise? First, our informed prior on α tends to be conservative relative to a diffuse prior. Bayesian updating shrinks posterior beliefs away from maximum likelihood estimates and towards the prior mean; in our application, small values for q tend to anchor beliefs at $\underline{\alpha}$ and make this shrinkage very powerful. Since our prior mean is negative, it is difficult for a manager with a short history of returns to demonstrate sufficient skill to raise his posterior mean above zero. This seems consistent with intuition and appropriate for this application. Second, despite several additional conservative assumptions, the empirical results for this sample of fund managers show that prior beliefs must be exceptionally strong in order to justify a zero-investment strategy in active managers – at least when the three-factor model is taken as the benchmark. The best performing manager has a positive value for $\tilde{\alpha}_{\min}$ when $q(25)$ is 0.00001, and ten managers exceed this threshold when $q(25)$ is 0.00056. In a frequentist test, such small values of $q(25)$ would be statistically indistinguishable from $q(25) = 0$, even in unbiased samples far larger than can currently be constructed. Thus, we doubt that zero investment in active managers can be justified on the basis of statistical evidence alone.

A natural next step would be to quantify the expected losses to an investor who ignored active managers in favor of the benchmarks. To carry out this analysis, one would need a full predictive distribution for both benchmark and manager returns (as discussed in Section II.E.), a utility function, and a complete description of the trading environment.²²

While a complete treatment of this portfolio-choice problem is outside the scope of our paper, simple calculations suggest that even relatively low levels of $q(25)$ might lead to significant investment in active managers for all but the most risk-averse investors. Over the range of $q(25)$ considered in Figure 5, there are many managers with positive values of $\tilde{\alpha}_{\min}$ and posterior standard deviations (for α) and residual standard deviations (σ) of similar magnitudes. Essentially, the trade-off comes down to a maximum expected loss of fees and costs if the best managers turn out to be unskilled versus the potential expected gain of several times this amount if the manager is skilled.

Our analysis assumed that *cost* was known and equal to six bp per month for all managers. How sensitive are our results to this assumption? Suppose for example, that we had used 9 bp as our baseline level. Then, for any given level of $q(25)$, $\underline{\alpha}$ (in Figure 1) shifts to the left, but both q and σ_{α} would increase (in order to maintain the same probability that α is greater than 25 bp). This effect can be seen by comparing Panels A and B in Table I. For the best performing funds, the second effect tends to dominate and $\tilde{\alpha}_{\min}$ becomes higher over most of the range of Figure 5. For example, the $\tilde{\alpha}_{\min}$ of Oakmark first becomes positive at $q(25) = 0.0007$ when *cost* is 9 bp. In fact, the most conservative possible results for the

²² See Pastor (1999) and Pastor and Stambaugh (1999b) for examples of this quantification.

best performing funds occur if we assume *cost* is zero: in this case, the $\tilde{\alpha}_{\min}$ of Oakmark first becomes positive at $q(25) = 0.0025$. Even in this unrealistic case, however, the results are qualitatively similar to the 6 bp case: for $cost = 0$ bp, both the highest and tenth highest $\tilde{\alpha}_{\min}(q(25))$ curves are only about 4 bp lower than they are in Figure 5. If *cost* (or *fee* or *a*) is uncertain, then we cannot obtain analytical solutions but our intuition is that these effects would be second-order compared to shifting the baseline level. For example, uncertainty around 6 bp should not have a larger effect than the most conservative possible shift to 0 bp. Different assumptions about *cost* would, however, affect inference about poorly performing funds, but this is not the main subject of our analysis.

A final point to consider is the incorporation of uncertainty into prior beliefs over q and σ_{α} . In principle, investigators can place any distribution they want on these prior parameters; solutions for the moments of the posterior distribution can then be calculated using numerical methods. It is also possible to handle uncertainty in the answers to the prior-elicitation questions of Section II.C. Uncertainty about $q(25)$ and/or $q(10)$ would translate directly into uncertainty about q and σ_{α} , and once again numerical methods can be used to solve for the posterior mean.

IV. Conclusion

In this paper, we develop a Bayesian procedure for evaluating investment managers. Using a flexible set of prior beliefs, we solve analytically for the posterior expectation of alpha. Our solution is expressed as a formula whose inputs are nothing more than modified moments of well-known distributions. This formula can be applied to managers across a wide

variety of contexts: mutual funds, hedge funds, brokerage analysts etc.; one can apply it to a single manager in isolation and it does not require a comprehensive or bias-free database. We also show how prior beliefs can be elicited by intuitive questions such as “what is the probability that a manager has an expected alpha greater than 25 bp per month?” and how the answers to these questions can be mapped into the parameters of the prior belief distribution.

We believe that this Bayesian approach provides a useful addition to the performance evaluation toolbox. As an illustration of our technique, we analyze the returns for the 1437 mutual-fund managers active at the end of 1996. Using the three-factor model of Fama and French (1993), we calculate the posterior expectation of alpha for each manager over a wide range of prior beliefs. To carry out this analysis, we impose the constraint that the average manager will have an alpha of zero *before* fees and transactions costs. Under this constraint, we ask “what prior beliefs would imply zero investment in active managers?” To justify such a zero-investment strategy, we find that a mean-variance investor would need to believe that less than 1 out of every 100,000 managers has an expected alpha greater than 25 bp per month. Furthermore, even slightly less conservative beliefs lead to positive and economically significant posterior expectations of alpha for many managers. Overall, our analysis suggests that even if the *average* manager is expected to underperform passive benchmarks, it requires very strong prior beliefs to imply zero investment in managers with the *best* past performance.

Appendix

A. Details for Section II.D.

Throughout this Appendix, we assume an informative prior on σ^2 , and, as in the text, a diffuse (improper) prior on β :

$$p(\beta) \propto 1 \quad (33)$$

$$p(\sigma^2) \propto \frac{1}{\sigma^{\nu_0+2}} \exp\left\{-\frac{h_0}{2\sigma^2}\right\} \quad (34)$$

Results in the text can be obtained by substituting $\nu_0 = 0$ and $h_0 = 0$ into the expressions below. Otherwise, the setup is the same as in Section II, and we make use of the same notation.

The likelihood for factors, unspecified in the text, is assumed to take the following form:

$$p(F_t | \mu_F, \Sigma_F) = N(\mu_F, \Sigma_F) \quad (35)$$

with realizations independent across t . The prior on μ_F and Σ_F is assumed to be diffuse:

$$p(\mu_F, \Sigma_F) \propto |\Sigma_F|^{-\frac{K+1}{2}} \quad (36)$$

This assumption is made for notational convenience and, except for the discussion of the predictive distribution below, does not affect our results.

Derivation of the posterior distribution for α :

From Bayes' rule, the joint posterior for $(\theta, \sigma^2, \mu_F, \Sigma_F)$ is given by:

$$\begin{aligned} p(\theta, \sigma^2, \mu_F, \Sigma_F | r, F) &\propto p(r | \theta, \sigma^2, F) p(F | \mu_F, \Sigma_F) p(\theta, \sigma^2) p(\mu_F, \Sigma_F) \\ &\propto p(\theta, \sigma^2 | r, F) p(\mu_F, \Sigma_F | F) \end{aligned} \quad (37)$$

where we have used the prior independence of (θ, σ^2) and (μ_F, Σ_F) , and the fact that the likelihood for r conditional on factors depends only on θ and σ^2 , while the likelihood for the factors depends only on (μ_F, Σ_F) . Therefore, (θ, σ^2) , and (μ_F, Σ_F) are independent in the posterior.

The above calculation implies

$$p(\theta, \sigma^2 | r, F) \propto p(r | \theta, \sigma^2, F)p(\theta, \sigma^2) \quad (38)$$

Because Z is independent from (μ_F, Σ_F) in the prior, and because the likelihood for factors does not depend on Z , the equations above are also valid conditional on a value of Z . We will make use of (38) throughout Appendix B.

The likelihood for r conditional on factors is given by:

$$\begin{aligned} p(r | \theta, \sigma^2, F) &\propto \frac{1}{\sigma^T} \exp \left\{ -\frac{1}{2\sigma^2} (r - X\theta)'(r - X\theta) \right\} \\ &= \frac{1}{\sigma^T} \exp \left\{ -\frac{1}{2\sigma^2} \left(S + (\theta - \hat{\theta})'X'X(\theta - \hat{\theta}) \right) \right\} \end{aligned} \quad (39)$$

Combining the likelihood and the prior yields:

$$\begin{aligned} p(\theta, \sigma^2 | Z = 1, r, F) &\propto \frac{1}{\sigma^{\nu_0+T+2}} \frac{1}{\sigma} \\ &\exp \left\{ -\frac{1}{2\sigma^2} \left(h_0 + S + \frac{s^2}{\sigma_\alpha^2} (\alpha - \underline{\alpha})^2 + (\theta - \hat{\theta})'(X'X)(\theta - \hat{\theta}) \right) \right\} \mathbf{1}_{\alpha > \underline{\alpha}} \end{aligned} \quad (40)$$

Integrating with respect to β yields:

$$p(\alpha, \sigma^2 | Z = 1, r, F) \propto \frac{1}{\sigma^{\nu_0+T-K+2}} \frac{1}{\sigma} \exp \left\{ -\frac{1}{2\sigma^2} \left(h_0 + S + \frac{s^2}{\sigma_\alpha^2} (\alpha - \underline{\alpha})^2 + \frac{(\alpha - \hat{\alpha})^2}{m} \right) \right\} \mathbf{1}_{\alpha > \underline{\alpha}} \quad (41)$$

Completing the square in α yields:

$$p(\alpha, \sigma^2 | Z = 1, r, F) \propto \frac{1}{\sigma^{\nu+2}} \frac{1}{\sigma} \exp \left\{ -\frac{1}{2\sigma^2} \left(h + (\lambda m)^{-1} (\alpha - \alpha')^2 \right) \right\} \mathbf{1}_{\alpha > \underline{\alpha}} \quad (42)$$

where

$$h = h_0 + S + \left(\frac{\sigma_{\alpha}^2}{s^2} + m \right)^{-1} (\hat{\alpha} - \underline{\alpha})^2 \quad (43)$$

$$\nu = \nu_0 + T - K + 2$$

Conditional on σ^2 , all terms involving σ^2 (as well as all terms involving only the data, such as $\hat{\alpha}$ and S), can be considered constants. Therefore,

$$p(\alpha | \sigma^2, Z = 1, r, F) \propto \exp \left\{ -\frac{1}{2\sigma^2} (\alpha - \alpha')^2 \right\} \mathbf{1}_{\alpha > \underline{\alpha}} \quad (44)$$

and we have shown (18).

In spite of the truncation, the functional form of the posterior is that of a conjugate prior distribution. Therefore, σ^2 can be integrated out of (42) using the properties of the gamma distribution (see, e.g., Gelman et al. (1995)). Define a change of variables $u = \frac{1}{2} A \sigma^{-2}$, where $A = h + (\lambda m)^{-1} (\alpha - \alpha')^2$. The resulting function of u is the pdf of a gamma distribution without the normalizing constant:

$$p(\alpha | Z = 1, r, F) \propto \left(\frac{A}{2} \right)^{-\frac{\nu+1}{2}} \left(\int u^{\frac{\nu-1}{2}} \exp\{-u\} du \right) \mathbf{1}_{\alpha > \underline{\alpha}} \quad (45)$$

Using the proportionality constant for the gamma distribution, and dividing through by h yields:

$$p(\alpha | Z = 1, r, F) \propto \left(\frac{h}{2} \right)^{-\frac{\nu+1}{2}} \Gamma \left(\frac{\nu+1}{2} \right) \left(1 + \frac{(\lambda m)^{-1} (\alpha - \alpha')^2}{\nu(h/\nu)} \right)^{-\frac{\nu+1}{2}} \mathbf{1}_{\alpha > \underline{\alpha}} \quad (46)$$

which is proportional to the pdf of a t -distribution. Therefore,

$$p(\alpha \mid Z = 1, r, F) \propto t_\nu \left(\alpha; \alpha', \frac{\lambda m h}{\nu} \right) \mathbf{1}_{\alpha > \underline{\alpha}} \quad (47)$$

Derivation of the posterior expectation $\tilde{\alpha}$ conditional on skill:

Because $p(\alpha \mid Z = 1, r, F)$ must integrate to 1:

$$p(\alpha \mid Z = 1, r, F) = \frac{1}{\int_{\underline{\alpha}}^{\infty} t_\nu \left(\alpha; \alpha', \frac{\lambda m h}{\nu} \right) d\alpha} t_\nu \left(\alpha; \alpha', \frac{\lambda m h}{\nu} \right) \mathbf{1}_{\alpha > \underline{\alpha}} \quad (48)$$

Let $\tilde{t} = \int_{\underline{\alpha}}^{\infty} t_\nu \left(\alpha; \alpha', \frac{\lambda m h}{\nu} \right) d\alpha$. Then

$$\begin{aligned} E[\alpha \mid Z = 1, r, F] &= \frac{1}{\tilde{t}} \frac{\Gamma(\nu + 1)/2}{\Gamma(\nu/2)} \sqrt{\frac{\lambda m h}{2\pi}} \int_{\underline{\alpha}}^{\infty} \alpha \left(1 + \frac{(\alpha - \alpha')^2}{\lambda m h} \right)^{-\frac{\nu+1}{2}} d\alpha \quad (49) \\ &= \alpha' + \frac{1}{\tilde{t}} \frac{\Gamma(\nu + 1)/2}{\Gamma(\nu/2)} \sqrt{\frac{\lambda m h}{2\pi}} \left(1 + \frac{(\underline{\alpha} - \alpha')^2}{\lambda m h} \right)^{-\frac{\nu-1}{2}} \\ &= \alpha' + \frac{\lambda m h}{\nu - 2} t_{\nu-2} \left(\underline{\alpha}; \alpha', \frac{\lambda m h}{\nu - 2} \right) \frac{1}{\int_{\underline{\alpha}}^{\infty} t_\nu \left(\alpha; \alpha', \frac{\lambda m h}{\nu} \right) d\alpha} \end{aligned}$$

The first line follows from the pdf of the t distribution, the second from a change of variables and the last from multiplying and dividing by the necessary constants.

Derivation of the likelihood ratio B :

Under our assumptions, $p(\theta, \sigma^2 \mid Z = 1, F) = p(\theta, \sigma^2 \mid Z = 1)$. Therefore,

$$p(r \mid Z = 1, F) = \int p(r \mid Z = 1, F, \theta, \sigma^2) p(\theta, \sigma^2 \mid Z = 1) d\theta d\sigma^2 \quad (50)$$

Substituting in for the likelihood and the prior and integrating with respect to β yields:

$$p(r \mid Z = 1, F) = C \sqrt{\frac{2}{\pi}} \frac{s}{\sigma_\alpha} \int_0^\infty \int_{\underline{\alpha}}^\infty \frac{1}{\sigma^{\nu+2}} \frac{1}{\sigma} \exp \left\{ -\frac{1}{2\sigma^2} \left(h_0 + S + \frac{s^2}{\sigma_\alpha^2} (\alpha - \underline{\alpha})^2 + \frac{(\alpha - \hat{\alpha})^2}{m} \right) \right\} d\alpha d\sigma^2 \quad (51)$$

where C is a constant that is identical for $Z = 1$ and $Z = 0$. Completing the square in α yields:

$$p(r | Z = 1, F) = C \sqrt{\frac{2}{\pi}} \frac{s}{\sigma_\alpha} \int \frac{1}{\sigma^{\nu+2}} \frac{1}{\sigma} \exp \left\{ -\frac{1}{2\sigma^2} \left(h + (\lambda m)^{-1} (\alpha - \alpha')^2 \right) \right\} d\sigma^2 \quad (52)$$

The form of the equation is the same as in (42). As above, the properties of the gamma distribution are used to integrate out σ^2 :

$$p(r | Z = 1, F) = C \sqrt{\frac{2}{\pi}} \frac{s}{\sigma_\alpha} \left(\frac{h}{2} \right)^{-\frac{\nu+1}{2}} \Gamma \left(\frac{\nu+1}{2} \right) \int_{\underline{\alpha}}^{\infty} \left(1 + \frac{(\lambda m)^{-1} (\alpha - \alpha')^2}{\nu(h/\nu)} \right)^{-\frac{\nu+1}{2}} d\alpha \quad (53)$$

The term inside the integral is proportional to a t -distribution. Therefore,

$$p(r | Z = 1, F) = C \left(\frac{\lambda m}{\sigma_\alpha^2/s^2} \right)^{\frac{1}{2}} \left(\frac{h}{2} \right)^{-\frac{\nu}{2}} \Gamma \left(\frac{\nu}{2} \right) 2 \int_{\underline{\alpha}}^{\infty} t_\nu(\alpha; \alpha', \lambda m h/\nu) d\alpha \quad (54)$$

The calculation for $Z = 0$ follows along the same lines. As above, after integrating out with respect to β , we obtain:

$$p(r | Z = 0, F) = C \int \frac{1}{\sigma^{\nu+2}} \exp \left\{ -\frac{1}{2\sigma^2} \left(S + h_0 + \frac{(\hat{\alpha} - \underline{\alpha})^2}{m} \right) \right\} d\sigma^2 \quad (55)$$

Integrating with respect to σ^2 yields:

$$p(r | Z = 0, F) = C \left(\frac{S + h_0 + (\hat{\alpha} - \underline{\alpha})^2/m}{2} \right)^{-\frac{\nu}{2}} \Gamma \left(\frac{\nu}{2} \right) \quad (56)$$

Therefore,

$$B = \left(\frac{\lambda m}{\sigma_\alpha^2/s^2} \right)^{\frac{1}{2}} \left(\frac{S + h_0 + (\sigma_\alpha^2/s^2 + m)^{-1} (\hat{\alpha} - \underline{\alpha})^2}{S + h_0 + m^{-1} (\hat{\alpha} - \underline{\alpha})^2} \right)^{-\frac{\nu}{2}} 2 \int_{\underline{\alpha}}^{\infty} t_\nu \left(\alpha; \alpha', \frac{\lambda m h}{\nu} \right) d\alpha \quad (57)$$

Both the numerator and the denominator are proportional to a t -distribution with $\nu - 1$ degrees of freedom (in one case, α was known, while in the other case we integrated with respect to α). Multiplying and dividing by a constant yields the expression in the text.

B. Diffuse Prior on α

In this section, we assume prior beliefs are given by:

$$p(\theta, \sigma^2) \propto \frac{1}{\sigma^2} \quad (58)$$

The likelihood is given by (39). Multiplying the prior and likelihood leads to the following simple form for the posterior:

$$p(\theta, \sigma^2 | r, F) \propto \frac{1}{\sigma^{T+2}} \exp \left\{ -\frac{1}{2\sigma^2} \left(S + (\theta - \hat{\theta})' X' X (\theta - \hat{\theta}) \right) \right\} \quad (59)$$

The posterior for θ conditional on σ^2 can be obtained by taking all terms involving σ^2 as constants:

$$p(\theta | \sigma^2, r, F) \propto \exp \left\{ -\frac{1}{2\sigma^2} (\theta - \hat{\theta})' X' X (\theta - \hat{\theta}) \right\} \quad (60)$$

That is, conditional on σ^2 , the posterior for θ is proportional to a normal distribution.

Therefore,

$$E[\alpha | \sigma^2, r, F] = \hat{\alpha} \quad (61)$$

By the law of iterated expectations,

$$E[\alpha | r, F] = \hat{\alpha} \quad (62)$$

Therefore, the posterior mean when the prior is diffuse is given by the maximum likelihood estimate.

C. Drawing from the predictive distribution

This section describes how to draw from the joint predictive distribution for r and F . We first show how to draw from posterior distribution of $(\alpha, \beta, \sigma^2)$. This is done by first drawing from the skilled distribution, then the unskilled distribution, and then the full posterior. Finally, we draw from the predictive distribution of r and F conditional on all parameters.

Drawing from the distribution conditional on $Z = 1$ (“skilled posterior”):

Consider random variables $\check{\sigma}^2$ and $\check{\alpha}$ such that:

$$\check{\sigma}^2 \sim IG\left(\frac{\nu}{2}, \frac{h}{2}\right) \quad (63)$$

$$\check{\alpha} | \sigma^2 = \check{\sigma}^2 \sim N(\alpha', \sigma'^2) \quad (64)$$

Then

$$p(\check{\alpha}, \check{\sigma}^2) \propto \frac{1}{\check{\sigma}^{\nu+2}} \frac{1}{\check{\sigma}} \exp\left\{-\frac{1}{2\check{\sigma}^2}(h + (\lambda m)^{-1}(\check{\alpha} - \alpha')^2)\right\} \quad (65)$$

Using the procedure above, but discarding the draw whenever $\check{\alpha} < \underline{\alpha}$ produces a distribution that has zero mass when $\check{\alpha} < \underline{\alpha}$, but where the relative densities of any other points is the same as in (65). By (42), this is exactly the joint distribution of α and σ^2 . Therefore, drawing $\check{\sigma}^2$ from $IG(\nu/2, h/2)$ and $\alpha | \sigma^2 = \check{\sigma}^2$ from $N(\alpha', \sigma'^2)$, and discarding the draws whenever $\check{\alpha} < \underline{\alpha}$, produces a draw from the joint posterior.

The posterior for β conditional on α and σ^2 follows from the properties of the multivariate normal:

$$\beta | \alpha, \sigma^2 \sim N(\hat{\beta} + m^{-1}y(\alpha - \hat{\alpha}), \sigma^2(Q - m^{-1}yy')) \quad (66)$$

where y and Q are submatrices of $(X'X)^{-1}$:

$$\begin{pmatrix} m & y' \\ y & Q \end{pmatrix} \quad (67)$$

Drawing from the distribution conditional on $Z = 0$ (“unskilled posterior”):

It follows from the likelihood and the prior, that conditional on $Z = 0$ we have

$$\sigma^2 \sim IG\left(\frac{\nu}{2}, \frac{h_0 + S + m^{-1}(\hat{\alpha} - \underline{\alpha})^2}{2}\right). \quad (68)$$

Conditional on σ^2 and on $\alpha = \underline{\alpha}$, β is drawn from (66).

Drawing from the full posterior:

For any given draw $(\alpha^j, \beta^j, \sigma^j)$, there is a probability \tilde{q} that the draw comes from the skilled posterior, and probability $1 - \tilde{q}$ that the draw comes from the unskilled posterior. Let u^j be a draw from the distribution with uniform mass on $[0, 1]$. If $u^j < \tilde{q}$, then $(\alpha^j, \beta^j, \sigma^j)$ is drawn from the skilled posterior. Otherwise, a draw is made from the unskilled posterior.

Drawing from the predictive distribution

From the likelihood:

$$p(r_{T+1} \mid \theta, \sigma^2, F_{T+1}) = N(\alpha + F_{T+1}\beta, \sigma^2) \quad (69)$$

and

$$p(F_{T+1} \mid \mu_F, \Sigma_F) = N(\mu_F, \Sigma_F) \quad (70)$$

The following procedure produces draws from the predictive distribution of r_{T+1} . First draw (μ_F, Σ_F) from their posterior distribution (see Pastor (1999)), then draw F_{T+1} from

(70), conditional on those values of μ_F and Σ_F . Next draw $(\alpha, \beta, \sigma^2)$ from the posterior, as described above. Finally, draw r_{T+1} from (69), conditional on those values of the parameters, and of F_{T+1} .

D. Derivation of the positive-investment condition for Section II.E.

In this Appendix we show that if an asset has a positive posterior expectation of α (i.e. $\tilde{\alpha} > 0$) then this asset will be held in positive quantities in the mean-variance tangency portfolio. Since our framework does not address the relation among multiple non-benchmark assets, we restrict ourselves to a universe of assets consisting of the benchmarks and a single manager with $\tilde{\alpha} > 0$. In a Bayesian setting the asset weights of the tangency portfolio are defined by

$$\omega = \frac{\tilde{V}^{-1}\tilde{E}}{\text{normalizing constant}} \quad (71)$$

where \tilde{E} and \tilde{V} are, respectively, the mean and variance covariance matrix of the predictive distribution of all the assets (see Pástor (1999)). Suppose that the first asset consists of the following portfolio:

- long the manager
- short the benchmark portfolios with weights equal to $\tilde{\beta} \equiv E[\beta \mid r, F]$.

We call this asset the “non-benchmark portfolio”. Since this portfolio is a linear combination of the manager and a benchmark portfolios, we drop the manager from the analysis and consider the remaining $K + 1$ portfolios: K benchmark portfolios and one non-benchmark portfolio. If we can show that the non-benchmark portfolio has a positive weight in the

tangency portfolio defined by (71), then it follows that the manager would have a positive weight as well.

Conditional on the parameters and past data, next period's total return on the non-benchmark portfolio is given by

$$R_{T+1} \equiv \alpha + (\beta - \tilde{\beta}) F_{T+1} + \sigma \varepsilon_{T+1}. \quad (72)$$

while next period's return on the factor portfolios is:

$$F_{T+1} \equiv \mu_F + \Sigma_F \eta_{T+1} \quad (73)$$

where $\varepsilon_{T+1} \sim N(0, 1)$ and $\eta_{T+1} \sim N(0, I_{K \times K})$ are independent. It follows from the law of iterated expectations that the posterior expectation of factor returns is given by:

$$\begin{aligned} E[F_{T+1} | r, F] &= E[E[\mu_F + \Sigma_F \eta_{T+1} | \mu_F, \Sigma_F, r, F] | r, F] \\ &= \tilde{\mu}_F \end{aligned} \quad (74)$$

The predictive expectation for the return on the non-benchmark portfolio is given by:

$$\begin{aligned} E[R_{T+1} | r, F] &= E[\alpha + (\beta - \tilde{\beta}) F_{T+1} + \varepsilon_{T+1} | r, F] \\ &= \tilde{\alpha} + E[(\beta - \tilde{\beta}) F_{T+1} | r, F] \end{aligned} \quad (75)$$

and, from the posterior independence of β and μ_F , we have

$$\begin{aligned} E[R_{T+1} | r, F] &= \tilde{\alpha} + E[E[(\beta - \tilde{\beta}) F_{T+1} | \theta, \sigma, \mu_F, \Sigma_F] | r, F] \\ &= \tilde{\alpha} + E[(\beta - \tilde{\beta}) \mu_F | r, F] \\ &= \tilde{\alpha} \end{aligned} \quad (76)$$

We next show that the predictive covariance between the returns to this benchmark portfolio and to the non-benchmark portfolio is zero. The predictive expectation of the product of the non-benchmark and benchmark returns is given by

$$\begin{aligned}
E[R_{T+1}F_{i,T+1} | r, F] &= E[F_{i,T+1}\alpha | r, F] + E[F_{i,T+1}(\beta - \tilde{\beta})F_{T+1} + \sigma\varepsilon_{T+1} | r, F] \quad (77) \\
&= E[\alpha E[F_{i,T+1} | \mu_F, \Sigma_F] | r, F] + \\
&\quad E[(\beta - \tilde{\beta})E[F_{i,T+1}F_{T+1} | \mu_F, \Sigma_F] | r, F] \\
&= \tilde{\alpha}\tilde{\mu}_F
\end{aligned}$$

The second line follows from the law of iterated expectations, and the third line follows from posterior independence (note that $E[F_{i,T+1}F_{T+1} | \mu_F, \Sigma_F]$ is a function of elements of μ_F and Σ_F). Therefore, the predictive covariance of F_{T+1} and $R_{i,T+1}$ is zero for all i . The structure of \tilde{V} is

$$\begin{pmatrix} v_{11} & 0 \\ 0 & V_{22} \end{pmatrix}, \quad (78)$$

where v_{11} is the predictive variance of the non-benchmark return and V_{22} is the predictive variance of the factors. Hence $\omega_1 = \tilde{\alpha}/v_{11}$, which is positive if and only if $\tilde{\alpha} > 0$.

References

- Barberis, Nicholas, 1999, Investing for the long run when returns are predictable, *Journal of Finance*, forthcoming.
- Bawa, Vijay S., Stephen J. Brown, and Roger W. Klein, 1979, *Estimation Risk and Optimal Portfolio Choice* (North-Holland, Amsterdam).
- Brown, Stephen J., 1979, Optimal portfolio choice under uncertainty: A Bayesian approach, in Vijay S. Bawa, Stephen J. Brown, and Roger W. Klein, Eds: *Estimation Risk and Optimal Portfolio Choice* (North-Holland, Amsterdam).
- Brown, Stephen J., and William N. Goetzmann, 1995, Performance persistence, *Journal of Finance* 50, 679-698.
- Carhart, Mark, 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57-82.
- Center for Research in Security Prices, CRSP Survivor Bias Free US Mutual Fund Data Base File Guide, 1999
- Chamberlain, Gary, 1998, Econometrics and Decision Theory, Working paper, Harvard University.
- Chevalier, Judith and Glenn Ellison, 1999, Are some mutual fund managers better than others? Cross-sectional patterns in behavior and performance, *Journal of Finance*, forthcoming.
- Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristic based benchmarks, *Journal of Finance* 52, 1035-1058.
- Daniel, Kent and Sheridan Titman, 1997, Evidence on the characteristics of cross-sectional variation in stock returns, *Journal of Finance* 52, 1-34.
- Fama, Eugene, 1970, Efficient capital markets: A review of theory and empirical work, *Journal of Finance* 25, 383-417.
- Fama, Eugene and Kenneth French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.
- Ferson, Wayne and Rudi Schadt, 1996, Measuring fund strategy and performance in changing economic conditions, *Journal of Finance* 51, 425-461.
- 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.

- Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin, 1995, *Bayesian Data Analysis*, Chapman and Hall.
- Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121-1152.
- Gruber, 1996, Another puzzle: The growth in actively managed mutual funds, *Journal of Finance* 51, 783-810.
- Hendricks, Darryl, Jayendu Patel and Richard Zeckhauser , 1993, Hot hands in mutual funds: The persistence of performance, 1974 - 1988, *Journal of Finance* 48, 93-130.
- Jensen, Michael C., 1968, The performance of mutual funds in the period 1945-1964, *The Journal of Finance* 23, 389-416.
- Jobson, J.D., and Bob Korkie, 1980, Estimation for Markowitz efficient portfolios, *Journal of the American Statistical Association* 75, 544-554.
- Jobson, J.D., Bob Korkie, and V. Ratti, 1979, Improved estimation for Markowitz efficient portfolios using James-Stein type estimators, *Proceedings of the American Statistical Association*, Business and Economics Statistics Section 41, 279-284.
- Jorion, Philippe, 1985, International portfolio diversification with estimation risk, *Journal of Business* 58, 259-278.
- Jorion, Philippe, 1986, Bayes-Stein estimation for portfolio analysis, *Journal of Financial and Quantitative Analysis* 21, 279-292.
- Jorion, Philippe, 1991, Bayesian and CAPM estimators of the means: Implications for portfolio selection, *Journal of Banking and Finance* 15, 717-728.
- Kandel, Shmuel, Robert McCulloch, and Robert F. Stambaugh, 1995, Bayesian inference and portfolio efficiency, *Review of Financial Studies* 8. 1-53.
- Kandel, Shmuel, and Robert F. Stambaugh, 1996, On the predictability of stock returns: An asset-allocation perspective, *Journal of Finance* 51, 385-424.
- Keim, Donald B., and Ananth Madhavan, 1997, Transactions costs and investment style: An inter-exchange analysis of institutional equity trades, *Journal of Financial Economics* 46, 265-92.
- Klein, Roger W., and Vijay S. Bawa, 1976, The effect of estimation risk on optimal portfolio choice, *Journal of Financial Economics* 3, 215-231.

- Lakonishok, Josef, Andrei Shleifer, and Robert Vishny, 1992, The structure and performance of the money management industry, *Brookings Papers: Microeconomics*, 339-391
- Lakonishok, Josef, Andrei Shleifer, and Robert Vishny, 1994, Contrarian investment, extrapolation, and risk, *Journal of Finance* 49, 1541-1578.
- McCulloch, Robert, and Peter E. Rossi, 1990, Posterior, predictive, and utility-based approaches to testing the arbitrage pricing theory, *Journal of Financial Economics* 28, 7-38.
- Malkiel, Burton, 1995, Returns from investing in mutual funds 1971 to 1991, *Journal of Finance* 50, 549-572.
- Pástor, Lubos, 1999, Portfolio selection and asset pricing models, *Journal of Finance*, Forthcoming.
- Pástor, Lubos, and Robert F. Stambaugh, 1999a, Costs of equity capital and model mispricing, *Journal of Finance* 54, 67-121.
- Pástor, Lubos, and Robert F. Stambaugh, 1999b, Comparing Asset-Pricing Models: An Investment Perspective, Working paper, The Wharton School.
- Perold, André F., 1988, The implementation shortfall: paper versus reality, *The Journal of Portfolio Management* 14, 4-9.
- Shanken, Jay, 1987, A Bayesian approach to testing portfolio efficiency, *Journal of Financial Economics* 19, 195-215.
- Shleifer, Andrei, 1999, *Clarendon Lectures: Inefficient Markets*, Oxford University Press, Forthcoming.
- Stambaugh, Robert F., 1997, Analyzing investments whose histories differ in length, *Journal of Financial Economics* 45, 285-331.
- Zheng, Lu, 1999, Is money smart? A study of mutual fund investors' fund selection ability, *Journal of Finance* 54, forthcoming.
- Zellner, Arnold, 1971, *An Introduction to Bayesian Inference in Econometrics*, John Wiley and Sons: New York.

</ref_section>

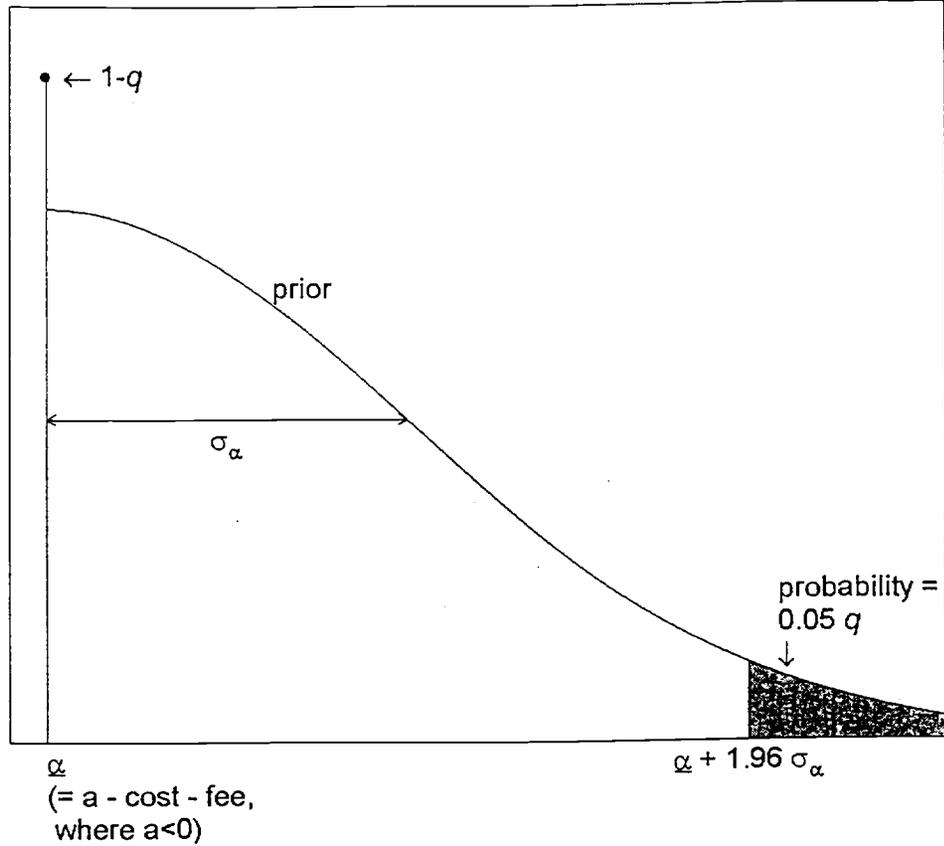


Figure 1:

Prior distribution of α

Figure 1 illustrates the (conditional) prior distribution for α . Please see Section I for a detailed discussion of the parameters in this figure. α is the intercept in a factor model (see equation (2)). q is the probability that a manager is skilled; conditional on skill, we have $\alpha \sim N(\underline{\alpha}, \sigma_\alpha^2)$ with a left truncation at $\underline{\alpha}$, where $\underline{\alpha} = a - \text{fee} - \text{cost}$ is the expected abnormal return for an unskilled manager, a is the expected negative return from transactions with skilled managers, and fee and cost are the manager's fees and transactions costs, respectively.

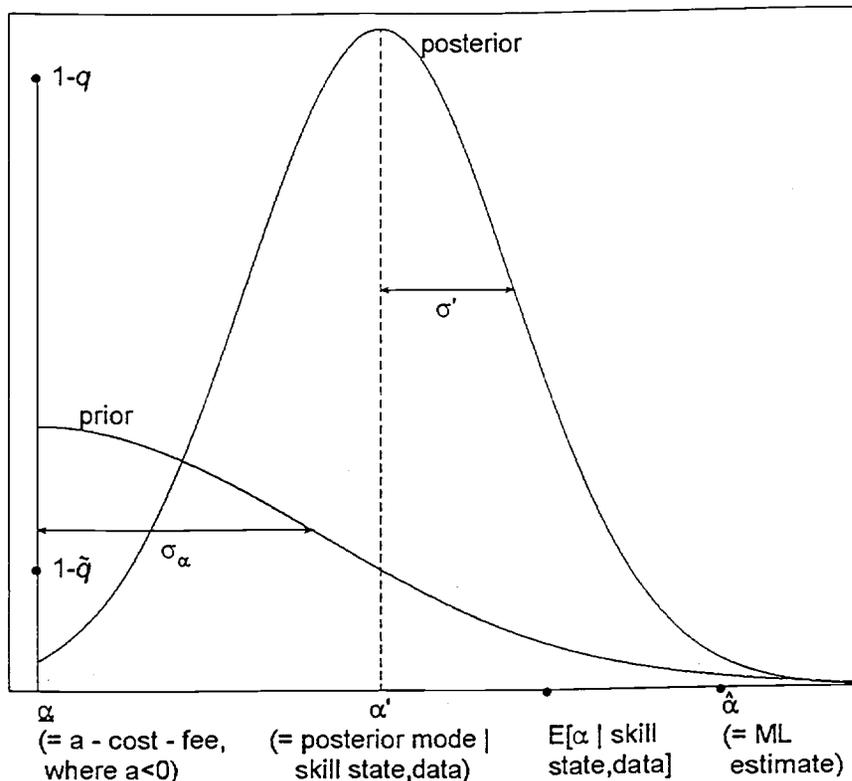


Figure 2:

Prior and posterior distribution of α

Figure 2 illustrates the prior and posterior distributions for α . α is the intercept in a factor model (see equation (2)). q is the probability that a manager is skilled; conditional on skill and $\sigma^2 = s^2$, we have $\alpha \sim N(\underline{\alpha}, \sigma_\alpha^2)$ with a left truncation at $\underline{\alpha}$, where $\underline{\alpha} = a - \text{fee} - \text{cost}$ is the expected abnormal return for an unskilled manager, a is the expected negative return from transactions with skilled managers, fee and cost are the manager's fees and transactions costs, respectively. Conditional on skill and $\sigma^2 = s^2$, the posterior distribution of α is $N(\alpha', \sigma'^2)$ with a left truncation at $\underline{\alpha}$. \tilde{q} is the posterior probability that $Z = 1$. $\hat{\alpha}$ is the maximum likelihood estimate of α .

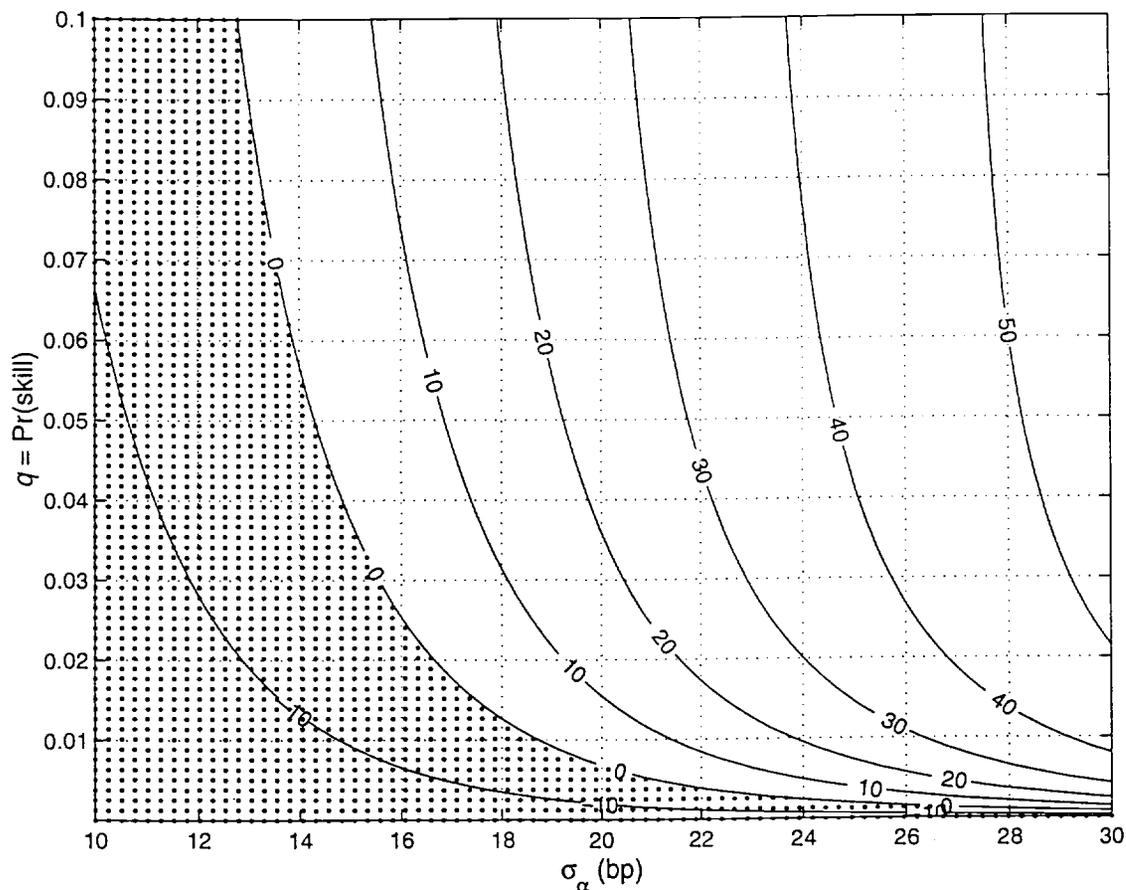


Figure 3:

Posterior expectation of α as a function of q and σ_α for the Oakmark Fund

Figure 3 shows level curves of $\bar{\alpha}$ (expressed in bp per month) in (q, σ_α) space. The shaded region indicates where $\bar{\alpha}$ is negative. α is the intercept in the Fama-French (1993) three-factor model (see equation (32)). $\bar{\alpha}$ is the posterior expectation of α . q is the probability that a manager is skilled; conditional on skill and $\sigma^2 = s^2$, we have $\alpha \sim N(\underline{\alpha}, \sigma_\alpha^2)$ with a left truncation at $\underline{\alpha}$, where $\underline{\alpha}$ is the expected abnormal return for an unskilled manager. The sample period for the Oakmark Fund is 9/91 - 12/96.

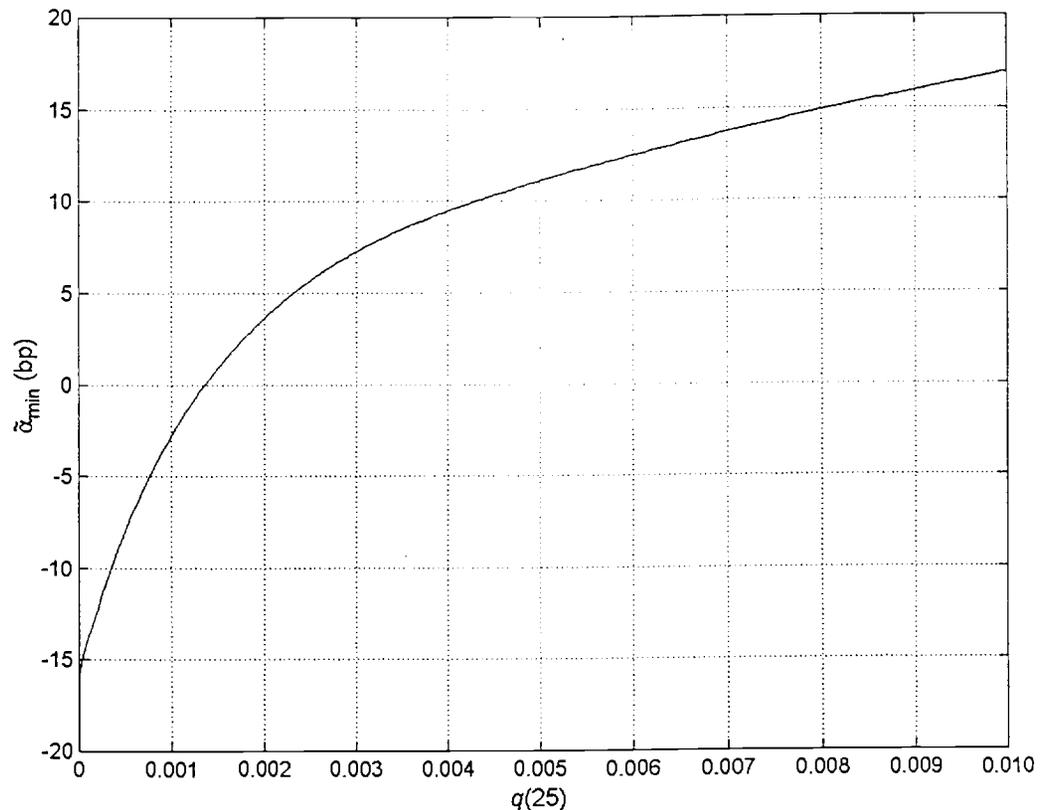


Figure 4:

$\tilde{\alpha}_{\min}$ as a function of $q(25)$ for the Oakmark Fund

Figure 4 shows the relation between $q(25)$ and $\tilde{\alpha}_{\min}$ for the Oakmark Fund. α is the intercept (bp per month) in the Fama-French (1993) three-factor model (see equation (32)). $\tilde{\alpha}$ is the posterior expectation of α . $\tilde{\alpha}_{\min}$ is defined as the minimum value of $\tilde{\alpha}$ for a fixed level of $q(25) \equiv P(\alpha > 25 \mid \sigma^2 = s^2)$. The prior constant s^2 is 0.00029. The sample period for the Oakmark Fund is 9/91 - 12/96.

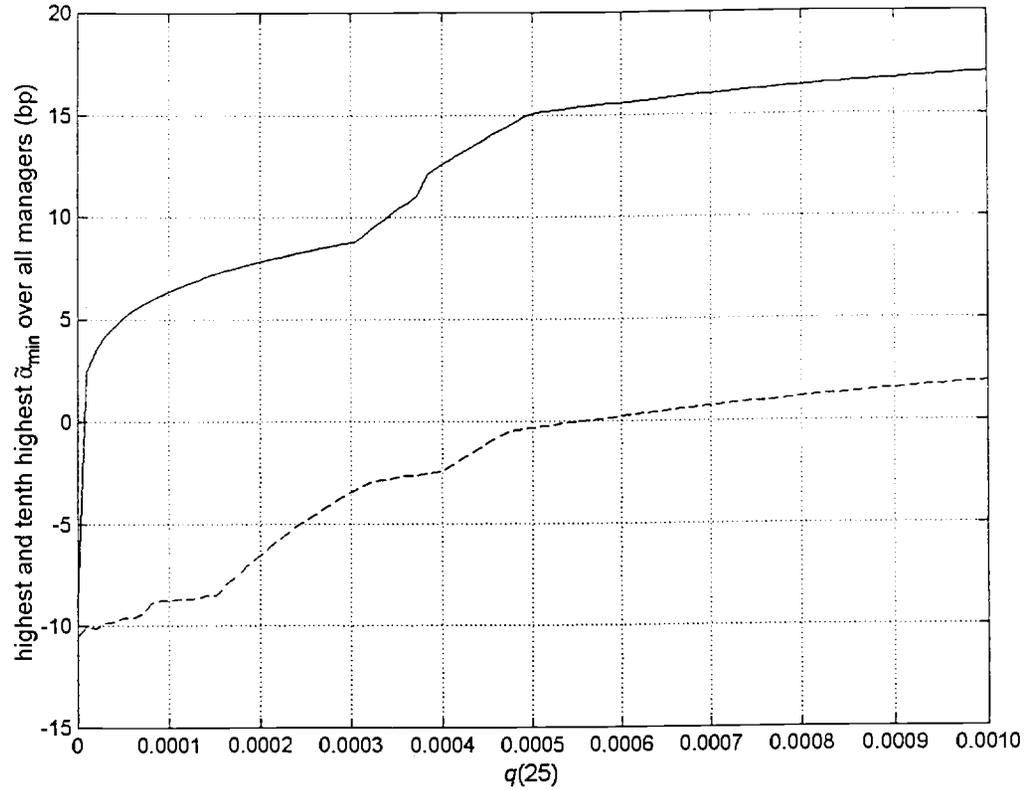


Figure 5:

The highest (solid) and tenth highest (dashed) $\tilde{\alpha}_{\min}$ over all managers

Figure 5 plots the highest and tenth highest $\tilde{\alpha}_{\min}$ over all 1437 managers for $q(25) \in (0, 0.001)$. α is the intercept (bp per month) in the Fama-French (1993) three-factor model (see equation (32)). $\tilde{\alpha}$ is the posterior expectation of α ; $\tilde{\alpha}_{\min}$ is defined as the minimum value of $\tilde{\alpha}$ for a fixed level of $q(25) \equiv \Pr(\alpha > 25 \mid \sigma^2 = s^2)$. The plot shows the highest and tenth highest value of $\tilde{\alpha}_{\min}$ among all managers for each level of $q(25)$. The prior constant s^2 is 0.00029. Figure 6 plots the same relationship for $q(25) \in (0.001, 0.01)$.

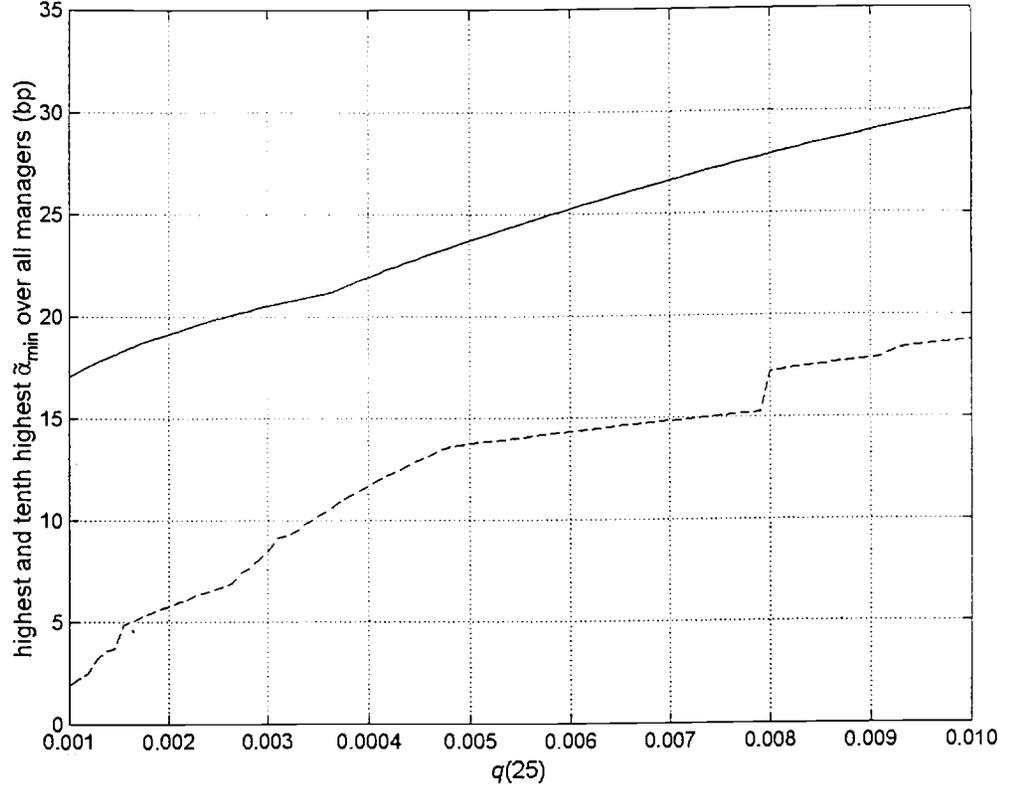


Figure 6:

The highest (solid) and tenth highest (dashed) $\tilde{\alpha}_{\min}$ over all managers

Figure 6 plots the highest and tenth highest $\tilde{\alpha}_{\min}$ over all 1437 managers for $q(25) \in (0.001, 0.01)$. α is the intercept (bp per month) in the Fama-French (1993) three-factor model (see equation (32)). $\tilde{\alpha}$ is the posterior expectation of α . $\tilde{\alpha}_{\min}$ is defined as the minimum value of $\tilde{\alpha}$ for a fixed level of $q(25) \equiv \Pr(\alpha > 25 \mid \sigma^2 = s^2)$. The plot shows the highest and tenth highest value of $\tilde{\alpha}_{\min}$ among all managers for each level of $q(25)$. The prior constant s^2 is 0.00029. Figure 5 plots the same relationship for $q(25) \in (0, 0.001)$.

Table I
Elicitation of Prior Beliefs

This table illustrates the mapping from $q(25)$, $q(10)$, fee and $cost$ into the prior parameters q , σ_α , a and $\underline{\alpha}$. $q(x)$ is defined as $P(\alpha > x \mid \sigma^2 = s^2)$; q is the probability that a manager is skilled; conditional on skill and $\sigma^2 = s^2$, we have $\alpha \sim N(\underline{\alpha}, \sigma_\alpha^2)$ with a left truncation at $\underline{\alpha}$, where $\underline{\alpha} = a - fee - cost$ is the expected abnormal return for an unskilled manager, a is the expected negative return from transactions with skilled managers, and fee and $cost$ are the manager's fees and transactions costs, respectively. σ_α , a and $\underline{\alpha}$ are expressed in bp per month; $q(25)$, $q(10)$ and q are expressed as probabilities.

Panel A: $fee = 8$ bp and $cost = 6$ bp					
$q(25)$	$q(10)$	q	σ_α	a	$\underline{\alpha}$
0.0001	0.0005	0.0024	19.22	-0.04	-14.04
	0.001	0.0082	15.61	-0.10	-14.10
	0.005	0.1601	11.84	-1.51	-15.51
0.001	0.005	0.0242	19.30	-0.37	-14.37
	0.01	0.0893	15.83	-1.12	-15.12
0.01	0.05	0.3301	20.50	-5.40	-19.40

Panel B: $fee = 8$ bp and $cost = 9$ bp					
$q(25)$	$q(10)$	q	σ_α	a	$\underline{\alpha}$
0.0001	0.0005	0.0029	19.94	-0.05	-17.05
	0.001	0.0106	16.24	-0.13	-17.14
	0.005	0.2950	12.54	-2.95	-19.95
0.001	0.005	0.0293	20.04	-0.47	-17.46
	0.01	0.1192	16.53	-1.57	-18.57
0.01	0.05	0.4691	21.78	-8.15	-25.15