

1 A Bayesian Approach to Estimating Mutual Fund Returns

Amir F. Atiya¹
Learning Systems Group
EE Department
Caltech, MC 136-93
Pasadena, CA 91125
amir@work.caltech.edu

Malik Magdon-Ismail
Learning Systems Group
EE Department
Caltech, MC 136-93
Pasadena, CA 91125
malik@work.caltech.edu

Accurate estimation of mutual fund statistics, such as mean return and standard deviation, can help shedding some light onto whether some mutual funds can consistently outperform the general market. In addition, it can help investors make more sound investment decisions. Because the available sample of returns is often not enough to give sufficient confidence in the estimates, we propose a novel Bayesian estimation approach, whereby the priors are obtained from the general market returns (by general market we mean the collection of thousands of available mutual funds). The justification for this is that any mutual fund is a subset of the general market, and it will therefore inherit some of its statistical properties. The advantages we gain is that we make use of the extensive sample size of the general market returns to fine tune our fund return estimates. The problem we face, however, is that the *a priori* density of the mean and standard deviation of the general market returns are unknown. The reason is that for an individual fund, the mean and standard deviation of the return can not be obtained exactly, as we can only estimate them from the finite sample available from historical data. In this paper we develop a new algorithm to tackle this estimation problem. We develop an EM-like algorithm that iteratively obtains the desired estimates. In addition, we present an approximate analytical solution to this problem. We use our approach to produce novel rankings of some available mutual funds. Our approach is not limited to mutual fund estimation and can be applied whenever parameters from a group of populations have to be estimated.

1.1 Introduction

There has been a raging debate about whether there are trading strategies or money managers that can consistently beat the general stock market. Although the dominant belief now is that the stock market exhibits some predictability, it is not clear whether this predictability offsets the associated risk. Researchers have looked at the performance of mutual funds for evidence. There are funds that consistently produce risk adjusted returns higher than the overall market. However, taking into account that there are over 4000 mutual funds, one cannot rule out the possibility that this is merely due to chance, by sheer number of funds. The topic of data snooping presents an analytical framework for assessing the effect and bias due to

1. Also holds a position as Research Scientist with Simplex Risk Management, Hong Kong

“model search” (see the studies in [White 1996, Lo and Mackinlay 1990]). Another approach to testing whether the differences in the returns of the funds is significant would be to perform an ANOVA analysis testing the hypothesis that the mean returns are different, [Casella and Berger 1990].

In this paper, we propose a new idea to estimate the mutual fund returns. The method we develop will shed some light into whether some funds can significantly and consistently beat the overall market. In addition, it will help investors have more accurate estimates of the mean and standard deviation of the funds’ returns, and therefore help them make more sound investment decisions. We will also produce a ranking of mutual funds according to several performance criteria using our approach.

1.2 Bayesian Estimation of Return Statistics

The Bayesian approach is widely studied by researchers in computational fields such as artificial intelligence, inference, pattern recognition, and neural networks (see [Bishop 1995] for example), but also has been applied in several finance problems. The usual criticism of most Bayesian approaches is that the choice of the prior tends to be arbitrary. In this study, however, we propose a more principled approach to obtaining the prior by using the density of the general market returns as our prior. The approach is suitable for our particular problem for several reasons. In the absence of compelling reasons, no funds will behave much better than the overall market, hence using the density of the overall market returns as our prior, acts as a suitable regularization factor, that will pull the return estimates justifiably towards those of the general market. Also, the general market is a superset of all mutual funds, and hence the individual funds will inherit some of the statistical properties of the general market, and this can be taken into account by using the Bayesian formulation we propose. Essentially, we view the funds as having been drawn from some distribution, i.e., there is some joint distribution for μ and σ^2 from which a particular fund’s mean and variance are drawn. We use the observed means and variances in the universe of funds to estimate this distribution. As an extreme example, consider a fund that has a very short track record of only a few months of very high returns. Although the sample mean and variance will give some indication of the fund’s return statistics, it will not be of a high confidence, and one would tend to assume the fund performance is closer to the market index performance than is apparent from simply looking at the sample mean.

Before we proceed to the details of our approach, we summarize the main diffi-

culty in applying our Bayesian technique. As already mentioned, it is the *a priori* density of the mean and standard deviation of the general market returns that is not known. If we consider estimating this density from the data, we face the problem that the data are noisy, the reason being that for the available funds, the sample means and standard deviations of the returns are not accurate, since they are only finite sample estimates from historical data. Thus, information about the *a priori* density that will be used to estimate the means and variances of the funds needs to be extracted from these noisy estimates of the means and variances. We are faced with an apparent dilemma: in order to estimate the mean and variance of a given mutual fund, we need the *a priori* density for the means and variances, and in order to estimate this density, we need estimates of the means and variances themselves. We propose an approach that overcomes this dilemma.

1.2.1 The Proposed Method

Let M_1, \dots, M_K denote the universe of mutual funds, and let μ_i and σ_i denote respectively the expectation and the standard deviation of the percentage returns of fund M_i . Assume we have available the historical returns of each fund (say the monthly returns): $x_i(t)$, $t = 1, \dots, N_i$ where N_i is the number of observations we have for fund i . Assume that the realizations $x_i(t)$ are independent over time and across funds², and let $X_i = \{x_i(1), \dots, x_i(N_i)\}$. Let us combine μ_i and σ_i into a single “parameter vector” $\alpha_i = (\mu_i, \sigma_i)^T$.

By Bayes rule,

$$p(\alpha_i|X_i) = \frac{p(X_i|\alpha_i)p(\alpha_i)}{p(X_i)} \quad (1.1)$$

where $p(\alpha_i)$ is the *a priori* probability density, meaning the density of the α parameter from which the α_i 's were drawn. It is this density that is to be estimated from the overall market. Having obtained $p(\alpha_i)$ and $p(X_i|\alpha_i)$, we are in principle done, because we can obtain $p(\alpha_i|X_i)$ using Bayes rule. We can now use $p(\alpha_i|X_i)$ to obtain our estimate of α_i according to any loss criterion we choose. We postpone the development of this more general method to a future presentation. Here, we will be satisfied with the mode of the density $p(\alpha_i|X_i)$, which can be obtained by finding the parameter estimate $\hat{\alpha}_i$ which maximizes the log of expression (1.1):

$$L_i = \ln(p(X_i|\alpha_i)) + \ln(p(\alpha_i)) - \ln(p(X_i)) \quad (1.2)$$

2. The independence assumption across funds can be relaxed considerably for the purposes of the algorithm we propose. This more general case will be discussed in a later presentation.

This is commonly known as the maximum *a posteriori* probability (MAP) estimate. We now consider each term in expression (1.2). One can write

$$p(X_i|\alpha_i) = \prod_{t=1}^{N_i} p(x_i(t)|\alpha_i) \quad (1.3)$$

which follows from our independence assumption for the daily returns. The term $p(x_i(t)|\alpha_i)$ can be considered Gaussian, because although it is well-known that the individual stock percentage returns are log-normal [Ritchken 1987], the overall average percentage return for a portfolio of several stocks can be approximated as Gaussian by virtue of the central limit theorem (for the case of continuous compounding, the individual stock returns themselves are Gaussian and so the Gaussian assumption would still be valid). Hence, we are justified in writing

$$p(x_i(t)|\alpha_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_i(t)-\mu_i)^2}{2\sigma_i^2}} \quad (1.4)$$

Now, consider $p(\alpha)$. The most self-evident approach would suggest estimating this density from the data, i.e. in terms of the μ_i 's and σ_i 's of the individual funds. Although we have plenty of data (up to 4000 mutual funds are available) to estimate $p(\alpha)$, these data are noisy since they give only estimates of each fund's μ_i and σ_i , rather than the true values. To alleviate this problem, we suggest the following: Assume that $p(\alpha)$ is determined with respect to the true α_j 's of the individual funds (which are unknown and not available), by a kernel density estimate:

$$p(\alpha) = \frac{1}{Kh^2} \sum_{j=1}^K \phi\left(\frac{\|\alpha - \alpha_j\|}{h}\right) \quad (1.5)$$

where h is the kernel bandwidth, and ϕ is the kernel, which is typically chosen as a Gaussian function³. For the two-dimensional case, the Gaussian kernel is given by

$$\phi(x) = \frac{1}{2\pi} e^{-\frac{1}{2}x^2} \quad (1.6)$$

Note that the density function $p(\alpha)$ is a function of the parameters α_j , that are exactly the quantities we are trying to estimate. This is one of the sources of

3. Note that in the asymptotic limit ($K \rightarrow \infty$), there is no loss of generality in the assumption that $p(\alpha)$ is given by (1.5) as long as $h \rightarrow 0$ and $Kh \rightarrow \infty$. This is because in this limit, $p(\alpha)$ in (1.5) will converge to the true prior [Silverman 1993].

difficulty that needs to be overcome. The term $p(X_i)$ can be obtained by integrating out the α_i parameter vector:

$$p(X_i) = \int p(X_i|\alpha_i)p(\alpha_i)d\alpha_i \quad (1.7)$$

Substituting from (1.3),(1.4),(1.5),(1.7) into (1.2), we get the following somewhat complicated expression for L_i

$$\begin{aligned} L_i = & -\ln(2\pi^{N_i/2}\sigma_i^{N_i}Kh^2) - \sum_{t=1}^{N_i} \frac{(x_i(t) - \mu_i)^2}{2\sigma_i^2} \\ & + \ln \left(\sum_{j=1}^K \phi \left(\frac{\sqrt{(\mu_i - \mu_j)^2 + (\sigma_i - \sigma_j)^2}}{h} \right) \right) \\ & - \ln \left(\int d\mu_i d\sigma_i \left[\prod_{t=1}^{N_i} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_i(t) - \mu_i)^2}{2\sigma_i^2}} \right] \right. \\ & \left. \frac{1}{Kh^2} \left[\sum_{j=1}^K \phi \left(\frac{\sqrt{(\mu_i - \mu_j)^2 + (\sigma_i - \sigma_j)^2}}{h} \right) \right] \right) \end{aligned} \quad (1.8)$$

Note that taking the exponential of (1.8), one obtains the posterior density $p(\alpha_i|X_i)$. We propose obtaining the estimates of the α_i 's as those that maximize the sum of the likelihood functions L_1, \dots, L_K , which corresponds to the MAP estimate for the $\alpha_1 \dots \alpha_K$ under the independence assumption across funds. The problem is that all the L_i are coupled together, thus we have an optimization problem in all the parameters $\alpha_1 \dots \alpha_K$. In principle, this is the optimization problem that we need to solve. In practice, however, we have 2 parameters for each fund, and, in all, about 4000 funds, giving a total of 8000 parameters. Thus, we are faced with a massive optimization problem, that is from the practical point of view intractable, even using some of the most sophisticated optimization procedures. To overcome this problem, we propose two approaches. The first one is an iterative procedure, while the second one is an analytic technique which attempts to undo the effect of the α_i 's being random rather than deterministic. The following are the two approaches:

1.2.2 An Iterative Estimation Procedure

In this procedure, we assume that we have a fairly reasonable starting choice of the parameters, say for the example the fund's sample mean and standard deviation, and hence we have a fairly reasonable estimate of $p(\alpha)$. Fixing $p(\alpha)$, we find the

optimal α_i 's. The L_i 's are coupled together only through $p(\alpha)$, thus, fixing $p(\alpha)$, we convert the coupled optimization problem into K two parameter optimization problems. Since the parameters have changed in the course of this iteration, an update for $p(\alpha)$ now needs to be performed, yielding a more accurate estimate of the *a priori* density. With the new $p(\alpha)$, we iterate again for the optimal α_i 's. We continue these iterations till convergence. We can see that the algorithm bears certain similarities to the EM algorithm [Dempster et al 1977]. The following are the details of the algorithm (we assume here Gaussian kernels for simplicity):

1 Initialize $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ to their sample values:

$$\hat{\mu}_i = \bar{x}_i \equiv \frac{1}{N_i} \sum_{t=1}^{N_i} x_i(t) \quad (1.9)$$

$$\hat{\sigma}_i^2 = s_i^2 \equiv \frac{1}{N_i} \sum_{t=1}^{N_i} (x_i(t) - \hat{\mu}_i)^2 \quad (1.10)$$

2 Calculate so-called β parameters:

$$\beta_{ij} = \frac{e^{-[(\hat{\mu}_i - \hat{\mu}_j)^2 + (\hat{\sigma}_i - \hat{\sigma}_j)^2]/2h^2}}{\sum_{k=1}^K e^{-[(\hat{\mu}_i - \hat{\mu}_k)^2 + (\hat{\sigma}_i - \hat{\sigma}_k)^2]/2h^2}} \quad (1.11)$$

Note that the matrix $\beta \equiv [\beta_{ij}]$ is a Markov matrix, meaning that it has nonnegative entries, and each row sums to 1. This property is useful in understanding the convergence properties of the algorithm.

3 For $i = 1$ to K , update the parameters as follows

$$\hat{\mu}_i(\text{new}) = \bar{x}_i + \text{bias}_i \quad (1.12)$$

where

$$\text{bias}_i = \hat{\mu}_i(\text{new}) - \bar{x}_i = \frac{1}{1 + \frac{N_i h^2}{\hat{\sigma}_i^2}} \sum_{j=1}^K \beta_{ij} (\hat{\mu}_j - \bar{x}_i) \quad (1.13)$$

where \bar{x}_i is the sample average, defined in Step 1. To obtain the update for $\hat{\sigma}_i$, first, solve the following one-dimensional polynomial equation by some search method:

$$\frac{y^4}{h^2} - \frac{y^3}{h^2} \sum_{j=1}^K \beta_{ij} \hat{\sigma}_j + N_i y^2 - N_i (s_i^2 + \text{bias}_i^2) = 0 \quad (1.14)$$

Suppose the solution to this equation occurs at y^* , then set

$$\hat{\sigma}_i(\text{new}) = y^* \quad (1.15)$$

(s_i^2 denotes the sample variance as defined in Step 1).

4 Repeat Steps 2 and 3 till convergence.

We briefly describe how we obtain the updates in step 3. Fixing $\hat{\mu}_j$ and $\hat{\sigma}_j$, the kernel centers in $p(\alpha_i)$, L_i becomes a function of μ_i and σ_i only:

$$L_i = -\ln \sigma_i^{N_i} - \sum_{t=1}^{N_i} \frac{(x_i(t) - \mu_i)^2}{2\sigma_i^2} + \ln \left(\sum_{j=1}^K e^{-\frac{(\mu_i - \hat{\mu}_j)^2 + (\sigma_i - \hat{\sigma}_j)^2}{2h^2}} \right) + \text{const.} \quad (1.16)$$

We now set

$$\frac{\partial L_i}{\partial \mu_i} = 0 \quad \text{and} \quad \frac{\partial L_i}{\partial \sigma_i} = 0,$$

Thus we have to solve a non-linear system of two equations in two unknowns. It is then easily seen that any fixed point of the iterative updates given by (1.13) and (1.14) represent a solution to this system. The interpretation of the mean update step of the algorithm is as follows. The new mean represents the weighted average of the sample mean and the weighted average of the mean returns of the other mutual funds (weighted according to how close the parameters are to the considered fund). This tells us an interesting fact that the funds having similar means and standard deviation have a more significant effect on the fund's return estimates. Such an interpretation for the standard deviation update is not as clear as that of the mean. We have proven that for the case of estimating the mean alone, the algorithm converges to a solution. We are still considering the general case of estimating both the mean and the standard deviation. In our experimental simulations, convergence always occurred.

1.2.3 An Approximate Analytic Solution

We present here an approximate analytic form for the solution which can either be used as a solution itself, as starting input to the initialization stage of the iterative algorithm above or as starting input to the full optimization problem. Suppose once again that

$$p(x_i(t)|\alpha_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_i(t) - \mu_i)^2}{2\sigma_i^2}} \quad (1.17)$$

Let \bar{x}_i , and s_i^2 denote the sample estimates for the mean and the variance. Then, it is well known that these two quantities are independent random variables and further that the distribution for the mean is normal and the distribution for the variance is $\chi_{N_i-1}^2$, which for large N_i is approximately normal [DeGroot 1989]. Thus we can write

$$\mu_i = \bar{x}_i + Z_i \quad \sigma_i^2 = \frac{N}{N-1} s_i^2 + Y_i \quad (1.18)$$

where Z_i and Y_i are independent normal random variables⁴ with $Z_i \sim N(0, \sigma_i^2/N_i)$ and $Y_i \sim N(0, \sigma_i^4/(N_i-1))$. We compute $p(\alpha)$ as follows:

$$p(\alpha) = \int dZ_i dY_i p(\alpha, Z_i, Y_i) = \int dZ_i dY_i p(\alpha|Z_i, Y_i) p(Z_i, Y_i) \quad (1.19)$$

Given Z_i, Y_i we know α_i from (1.18) because both \bar{x}_i and s_i^2 are computable from the data. Therefore using (1.5), we can obtain $p(\alpha|Z_i, Y_i)$ as follows⁵:

$$p(\alpha|Z_i, Y_i) = p(\mu, \sigma^2|Z_i, Y_i) = \frac{1}{K h^2} \sum_{i=1}^K \frac{1}{2\pi^2} e^{-\frac{[(\mu - \bar{x}_i - Z_i)^2 + (\sigma^2 - \frac{N}{N-1} s_i^2 - Y_i)^2]}{2h^2}} \quad (1.20)$$

where we are assuming a Gaussian kernel estimator. We use the fact that Z_i and Y_i are independent Normals to obtain $p(Z_i, Y_i)$ as follows:

$$p(Z_i, Y_i) = \frac{1}{\sqrt{2\pi\sigma_i^2/N_i}} e^{-\frac{Z_i^2}{2\sigma_i^2/N_i}} \frac{1}{\sqrt{2\pi\sigma_i^4/(N_i-1)}} e^{-\frac{Y_i^2}{2\sigma_i^4/(N_i-1)}} \quad (1.21)$$

Note that $p(Z_i, Y_i)$ depends on σ_i^2 . As we will be interested in an asymptotic expansion, we will eventually replace σ_i^2 by s_i^2 . Finally, we can use (1.20) and (1.21) in (1.19) and doing the integration, we find:

$$p(\mu, \sigma^2) = \frac{1}{K} \sum_{i=1}^K \frac{1}{h_{\bar{x}_i}} \tilde{\phi}\left(\frac{\|\mu - \bar{x}_i\|}{h_{\bar{\mu}_i}}\right) \frac{1}{h_{\hat{\sigma}_i^2}} \tilde{\phi}\left(\frac{\|\sigma^2 - s_i^2\|}{h_{\hat{\sigma}_i^2}}\right) \quad (1.22)$$

where $h_{\bar{\mu}_i}^2 \approx h^2 + s_i^2/(N_i-1)$ and $h_{\hat{\sigma}_i^2}^2 \approx h^2 + s_i^4/(N_i-1)$ and $\tilde{\phi}(x) = e^{-x^2/2}/\sqrt{2\pi}$. We see that this is formally identical to (1.5) with two (different) larger kernel widths corresponding to the mean and variance estimation. Thus we see that as far

4. We use the notation $X \sim N(a, b)$ to mean that X is a Normal random variable with mean a and variance b .

5. For the purposes of the present discussion, we consider the parameter vector as $\alpha_i = [\mu_i, \sigma_i^2]^T$.

as the density estimation is concerned, the fact that \bar{x}_i and s_i^2 are noisy versions of μ_i and σ_i^2 appears as increased regularization for the kernel estimate. To complete the analysis, we use (1.22) to obtain $p(\alpha_i)$ for the purposes of maximizing (1.2). From (1.2) we see that the prior introduces a correction to the maximum likelihood estimator that is $O(1/N_i)$. We thus compute the estimates of μ_i, σ_i^2 that maximize (1.2) as a series in $1/N_i$. This leads to the following expressions for $\hat{\mu}_i, \hat{\sigma}_i^2$, our estimates of μ_i, σ_i^2 :

$$\hat{\mu}_i = \bar{x}_i + \frac{a_i}{N_i} + O\left(\frac{1}{N_i^2}\right) \quad (1.23)$$

$$\hat{\sigma}_i^2 = s_i^2 + \frac{b_i}{N_i} + O\left(\frac{1}{N_i^2}\right) \quad (1.24)$$

where

$$a_i = s_i^2 \sum_{j=1}^K \beta_{ji} \frac{\bar{x}_j - \bar{x}_i}{h_{\hat{\mu}_j}^2} \quad (1.25)$$

$$b_i = s_i^4 \sum_{j=1}^K \beta_{ji} \frac{s_j^2 - s_i^2}{h_{\hat{\sigma}_j^2}^2} \quad (1.26)$$

where

$$\beta_{ij} = \frac{\lambda_{ij}}{\sum_{k=1}^K \lambda_{ki}}, \quad \lambda_{ij} = \frac{1}{h_{\hat{\mu}_i}} \tilde{\phi}\left(\frac{\|\bar{x}_j - \bar{x}_i\|}{h_{\hat{\mu}_i}}\right) \frac{1}{h_{\hat{\sigma}_i^2}} \tilde{\phi}\left(\frac{\|s_j^2 - s_i^2\|}{h_{\hat{\sigma}_i^2}}\right) \quad (1.27)$$

β_{ij} is analogous to the β -matrix of the previous section where h was used in place of $h_{\hat{\mu}_i}, h_{\hat{\sigma}_i^2}$. β_{ij} represents the interaction between the mutual funds represented by the prior. β_{ij} is significant only for mutual funds with $\hat{\alpha}_i \approx \hat{\alpha}_j$. We see that the prior affects our estimates for α_i only through this interaction term.

1.3 Simulations

We have applied both the iterative and analytical methods to a collection of 1527 mutual funds using up to five years of data (ending June 1998) for each fund. The purpose is to obtain a novel ranking of mutual funds, and compare this ranking with the conventional ranking. We rank the mutual funds according to their mean return, μ , and their Sharpe ratio, μ/σ .

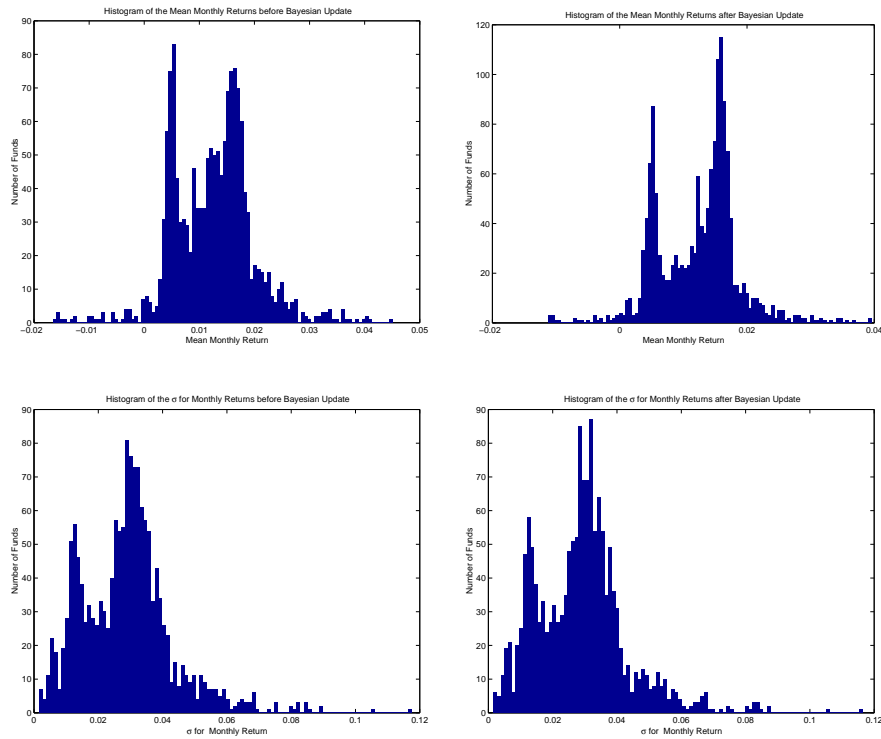


Figure 1.1

Histograms of the distributions of μ and σ before and after Bayesian update (using the analytical solution). The top two are for μ and the lower two are for σ . The graphs on the right show the histograms after update and the ones on the left are before update.

Tables 1.1, 1.2 and 1.3 show such rankings according to the sample values, the iterative updates and analytical solution respectively. One can see that for both proposed methods, as well as for the standard method of using the sample mean and the sample Sharpe ratio, the top ten funds are mostly the same. What is different is the relative positions of these top ten funds. We point out here that there is perhaps a slight bias in the Bayesian updating because in estimating the prior, one really wants to have access to all the funds that appeared, their means and variances. However, in the market, there is a selection bias that favors the survival of funds with higher μ 's and lower σ 's. To the extent that this is so, our prior will be biased in those directions with respect to the true prior. Thus even

Ranking using \bar{x}_i			Ranking using \bar{x}_i/s_i		
Fund	$\mu(\%)$	σ	Fund	$\mu(\%)$	μ/σ
TrnsAm Sml Co	69.67	0.054	State Str Glbl Adv Yld	4.91	2.56
State Str Aurora Fnd	62.54	0.047	Str Muni Adv Fnd	4.81	2.45
Frnkln Fin Svcs A	61.19	0.028	Str Adv Fnd	6.16	2.37
Frnkln Fin Svcs B	61.19	0.027	DFA One Yr Fxd Inc	5.31	1.83
PBHG Mid Cap	59.73	0.040	Frnkln Fin Svcs B	61.19	1.49
Oakmark Sel	57.12	0.046	Frnkln Fin Svcs A	61.19	1.47
Warburg Pncs Hlth	55.04	0.029	Str Sh Trm Glbl Bnd	8.32	1.36
TrnsAm Aggr Gr	54.52	0.055	Harbor Shrt Dur	5.43	1.29
Janus Sp Sit	53.33	0.037	Htckis & Wiley low Dur	8.46	1.28
Vista Grp Sml Cap B	53.33	0.050	UAM Trst/FPA Cresc	26.42	1.27

Table 1.1

Ranking of the funds according to sample annualized mean return and Sharpe ratio.

Ranking using $\hat{\mu}_i$ from Iter Proc			Ranking using $\hat{\mu}_i/\hat{\sigma}_i$ from Iter Proc		
Fund	$\mu(\%)$	σ	Fund	$\mu(\%)$	μ/σ
TrnsAm Sml Co	69.67	0.049	State Str Glbl Adv Yld	4.91	3.02
State Str Aurora Fnd	59.72	0.042	Str Muni Adv Fnd	4.81	2.61
Oakmark Select	59.66	0.042	Str Adv Fnd	6.11	2.44
PBHG Mid Cap	59.64	0.036	DFA One Yr Fxd Inc	5.28	1.95
Frnkln Fin Svcs A	59.36	0.025	Frnkln Fin Svcs A	59.36	1.56
Frnkln Fin Svcs B	59.36	0.025	Frnkln Fin Svcs B	59.36	1.56
Warburg Pncs Health	59.22	0.026	Warburg Pncs Health	59.22	1.55
TrnsAm Aggr Gr	51.00	0.049	Harbor Shrt Dur	4.49	1.43
Warburg Pncs Sml Co	51.00	0.049	Icon Shrt Trm Fxd Inc	3.70	1.38
Vista Grp Sml Cap B	51.00	0.049	Htckis & Wiler Low Dur	5.38	1.34

Table 1.2

Ranking of the funds according to annualized mean return and Sharpe ratio from the iterative procedure

these Bayesian estimates will be slightly optimistic - this, however, is not a flaw with the Bayesian estimate itself, because in principle we could have access to all the funds that ever traded. We merely note it here as a practical point to be kept in mind when applying the Bayesian method.

Figure 1.1 displays the effect that the Bayesian update of the parameter estimation has on the distribution of the μ 's and σ 's. Notice that the distributions

Ranking using $\hat{\mu}_i$ from Anal Sol			Ranking using $\hat{\mu}_i/\hat{\sigma}_i$ from Anal Sol		
Fund	$\mu(\%)$	σ	Fund	$\mu(\%)$	μ/σ
Frnkln Fin Svcs B	59.31	0.028	State Str Glbl Adv Yld	4.91	2.56
Frnkln Fin Svcs A	59.12	0.028	Str Muni Adv Fnd	4.82	2.45
Warburg Pncs Health	54.53	0.030	Str Adv Fnd	6.15	2.37
PBHG Mid Cap	52.52	0.040	DFA One Yr Fxd Inc	5.31	1.84
State Str Aurora Fnd	50.62	0.048	Frnkln Fin Svcs B	59.31	1.41
TrnsAm Sml Co	50.39	0.053	Frnkln Fin Svcs A	59.12	1.39
Oakmark Select	49.57	0.046	Str Sh Trm Glbl Bnd	4.49	1.35
Janus Sp Sit	49.34	0.037	Icon Shrt Trm Fxd Inc	8.82	1.34
FBR Sml Cap Fin	48.22	0.035	Harbor Shrt Dur	3.62	1.32
PBHG Sml Cap Val	45.88	0.040	Htkis & Wiley low Dur	5.42	1.29

Table 1.3

Ranking of the funds according to annualized mean return and Sharpe ratio from the analytic solution.

after Bayesian update are more concentrated indicating that the outliers have been brought closer to the general market (in a principled way).

1.4 Conclusions

We have developed a method for understanding the small sample effects that could lead to misleading estimates of a funds performance using a Bayesian approach. The Bayesian approach requires a prior and we have provided a natural prior that is available in the case where a large population of groups is available (in our case, each group consisted of a particular fund's returns). The general effect of using this prior is to pull the parameters of a particular group for which only a few observations are available toward the statistics of the overall population. Intuitively this seems like the correct thing to do and what we have presented is a quantitative way of doing so.

1.5 Acknowledgments

We would like to acknowledge the generous support of the National Science Foundation under NSF Cooperative Agreement EEC 9402726 to the Center for Neuro-morphic Systems Engineering.

References

- Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford.
- Casella, G. and R. L. Berger. 1990. *Statistical Inference*. Duxbury, Belmont, CA.
- DeGroot, M. H. 1989. *Probability and Statistics*. Addison-Wesley, Reading, Massachusetts.
- Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the E-M algorithm. *Journal of the Royal Statistical Society B*, 39:1-38.
- Lo, A. and A. Mackinlay. 1990. Data snooping biases in tests of financial asset pricing models. *Review of Financial Studies*, 3(3):431-467.
- Ritchken, P. 1987. *Options: Theory, Strategy and Applications*. Harper-Collins.
- Silverman, B. 1993. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, UK.
- White, H. 1996. Data snooping with confidence. In *Neural Networks in the Capital Markets*, Pasadena, CA.