Evaluating Mutual Fund Performance and its Persistence using Shrinkage Estimators *

Joop Huij† Marno Verbeek‡

June 30, 2003

Abstract

In this study, we evaluate the usefulness of shrinkage estimation in analyzing mutual fund performance and its persistence. Shrinkage estimators exploit information contained in the cross-section of mutual fund returns and enable a more accurate estimation of mutual fund alphas. We present and analyze three alternative shrinkage estimators and investigate their properties in a simulated sample of mutual funds in comparison with standard OLS estimators. We focus on the implications of using these estimators as an alternative approach to persistence analysis. All shrinkage estimators are fairly easy to compute using information from returns only. The results indicate that shrunk estimates are substantially more accurate than OLS in realistic settings. Consequently, persistence studies using shrunk estimates appear to be significantly more reliable.

Keywords: mutual fund performance; small sample bias; Stein-rule; iterative emperical Bayes; Gibbs sampling

*We are grateful to Georges Bresson and Alain Pirotte for helpful discussions.

†Dept. of Financial Management, Erasmus University Rotterdam, The Netherlands, jhuij@fbk.eur.nl.

‡Dept. of Financial Management and Econometric Institute, Erasmus University Rotterdam, The Netherlands, mverbeek@fbk.eur.nl.

1 Introduction

While the value of active portfolio management is often disputed, see e.g. Malkiel (1995), the performance measures on which this type of analysis is grounded are often based on short time series so that they are typically affected by small sample bias. Consequently, mutual fund rankings are largely based on luck rather than skills. Shrinkage estimators provide a simple way to reduce the impact of random luck components upon performance measures.

A common approach to determine mutual fund performance is via the intercept term in a time series regression of the fund's excess returns upon the excess returns of one or more benchmark factors. The resulting measure, typically referred to as alpha, measures the additional return that the fund is able to generate above a passive portfolio based on the benchmark factors. To analyze persistence in performance, researchers often distinguish two subperiods, a ranking period and a post-ranking or evaluation period, which may both be as short as 12 months. First, the performance of each individual fund is estimated over the ranking period. Next, funds are ranked, e.g. in deciles, on the basis of their estimated performance, and subsequently, the post-ranking performance in determined within each decile. This allows one to analyze the persistence in fund performance and to investigate the presence of so-called hot hands, see e.g. Hendricks et al. (1993). If winners repeat, active fund selection based upon historical performance is of interest to individual investors. This phenomenon also has implications for the efficient market hypothesis (EMH). However, the limited number of returns in each subperiod suggests that performance measures and fund rankings may be substantially affected by noise. Consequently, it may be fruitful to incorporate additional information in the estimation procedure to obtain more accurate performance measures.

In this paper we investigate to what extent standard ordinary-least-squares (OLS) estimates of performance measures are affected by noise and how these estimates are improved by using shrinkage estimation. Shrinkage estimators exploit the entire cross-section of mutual funds to estimate each individual fund's performance. Intuitively, a shrinkage estimator computes the weighted average of an individual time series estimate and the overall pooled estimate as both estimators contain relevant information. This way, part of the estimation error is eliminated and the accuracy of each individual estimate may improve. The use of shrinkage estimators in finance is not completely new. For example, Blume (1971) advocates a correction method to estimate beta coefficients for individual securities, while Vasicek (1973) proposes a simple Bayesian approach to improve the cross-section of estimated betas. A number of recent papers incorporates additional information in the estimation of mutual fund alphas using a Bayesian approach, see, for example, Baks et al. (1999), Pastor and Stambaugh (2001) and Busse and Irvine (2002). These approaches somehow require the researcher to formalize his prior beliefs or knowledge, based on, e.g., information on expense ratios, historical performance or non-benchmark assets.

The approaches in this paper do not require any information about the mutual funds in addition to their returns. The increase in estimation accuracy is entirely due to the exploitation of the fact that fund alphas tend to fluctuate around a common mean, close to zero. As a result, large positive alpha estimates are likely to be overestimated, while large negative alphas tend to be underestimated. Shrinking the estimators towards their common mean reduces this problem and improves accuracy.

Section 2 introduces three alternative shrinkage estimators. Section 3 presents the set-up of a Monte Carlo study to analyze the properties of the three shrinkage estimators in comparison with the usual OLS approach. The results of the simulation exercise are presented and discussed in Section 4, while Section 5 concludes.

2 Shrinkage Estimators

In this section we introduce three alternative approaches to shrinkage estimation. Typically, performance of mutual funds is measured through the intercept α_i in a regression of excess returns of a mutual fund upon excess returns of one or more benchmarks:

$$y_{it} = \alpha_i + x_{1t}\theta_{1i} + x_{2t}\theta_{2i} + \ldots + x_{kt}\theta_{ki} + \varepsilon_{it}, \tag{1}$$

where y_{it} denotes the excess return of fund i in period t, and i, x_j is the excess return on factor j. Stacking all T_i observations for fund i in a vector, we can write the above model in matrix notation as

$$y_i = X_i\theta_i + \varepsilon_i, \tag{2}$$

where X_i denotes the T_i by k matrix of excess returns of the benchmarks factors, and θ_i denotes a $k + 1$ by 1 vector of unknown parameters, containing α and the k factor sensitivities of fund i. A wide range of alternative choices can be made for the benchmarks. The most frequently used models are the one-factor CAPM, where the market portfolio provides the benchmark, the three-factor model from Fama and French (1992), and the four-factor from Carhart (1997) model. Fama and French add two factors based on firm size and the ratio of book-to-market value (the "value" factor), while Carhart adds a momentum factor. Alpha measures the additional return the mutual fund is able to generate above a passive benchmark portfolio based on the included factors.

Usually the above model is estimated by using ordinary-least-squares (OLS) using a given time series. An empirical problem with this approach is that time series are usually very limited so that only a small number of observations is available to estimate performance. As a result, estimates are typically inaccurate. By incorporating additional information performance estimators may become more accurate. Additional information can be obtained by exploiting the large cross-section of mutual funds that is available. One way to incorporate this information is by using shrinkage estimators.

To explain shrinkage estimation, consider the following situation: if the funds are all the same, the data can be pooled and each fund can be described by one overall estimate. In this situation, a pooled estimate is more accurate than a separate time series estimate, because it is based on more observations. However, if the funds have no similarity whatsoever, a pooled estimate is uninformative and estimates should be made separately for each fund. Now, if there is some similarity between the funds, both the pooled and the separate time series estimates contain information. With shrinkage estimation, the resulting estimate is a weighted average of both estimates. Another way of saying this is that the separate cross-section estimates are "shrunk" towards the overall pooled estimate. The degree of this shrinkage depends on the degree of similarity: the higher the degree of similarity, the more information the pooled estimate contains compared to the individual time series estimates, and the more cross-sectional estimates are shrunk towards the overall pooled estimate. A wide range of alternative approaches to shrinkage estimation exists. From a comprehensive overview of these estimators we refer to Maddala et al. (1997) and Blattberg and George (1991). In the remainder of this section, we will discuss the main approaches to shrinkage estimation.

2.1 Stein-rule

A popular approach to shrinkage estimation is the Stein-rule estimator. The criterium applied by the Stein-rule to weigh the individual estimates and the pooled estimate is the degree of homogeneity. The degree of homogeneity can be defined as the degree to which individual time series estimates are described by an overall pooled estimate. The higher the degree of homogeneity, the more information the pooled estimate contains as compared to the individual estimates, and the more individual estimates are shrunk towards the pooled estimate. The Stein-rule estimator $\tilde{\theta}_i$ is formulated as follows

$$\tilde{\theta}_i = (1 - \frac{c}{F})\hat{\theta}_i + \frac{c}{F}\hat{\theta}_p, \tag{3}$$

where $\hat{\theta}_i$ denotes the OLS estimator for the parameters of the ith cross-section, and $\hat{\theta}_p$ denotes the pooled OLS estimator. The coefficient c/F denotes a measure for homogeneity where the constant c is given by

$$c = \frac{(N-1)k - 2}{N(T-k) + 2}, \tag{4}$$

and F denotes the test statistic for the null hypothesis

$$H_0 : \theta_1 = \theta_2 = \ldots = \theta_N = \theta_p. \tag{5}$$

This test statistic is a criterion for falsifying the assumption of complete homogeneity of the parameters. The most common test for this hypothesis is a Chow test, which involves the test statistic

$$F = \frac{(e'e - e_1'e_1 - e_2'e_2 - \ldots - e_N'e_N)/(N-1)k}{(e_1'e_1 - e_2'e_2 - \ldots - e_N'e_N)/N(T-k)}, \tag{6}$$

where $e'e$ denotes the residual sums of squares of the pooled OLS estimator, and $e_i'e_i$ denotes the residual sums of squares of the OLS estimator of the ith cross-section.

2.2 Iterative empirical Bayes

Basically, the Stein-rule estimator tests to what degree parameters are described by an overall pooled estimate, and conditional upon this degree of homogeneity, parameters are shrunk towards the pooled estimate. In this approach, the degree of shrinkage is constant for all individual estimates. However, better results can probably be obtained if the degree of shrinkage is variable for individual estimates: some estimated values are "less likely" than others and should therefore be shrunk more towards the pooled estimate. To determine to how likely estimates are, it is most convenient to use a Bayesian framework. In this context, a prior distribution is specified for the unknown parameters, specifying the degree of prior uncertainty about θ_i. Next, a posterior distribution is derived, conditional upon the observed data. A common choice for the prior distribution is a normal distribution, given by

$$\theta_i \sim N(\mu, \Sigma). \tag{7}$$

Under the assumption of i.i.d. normal error terms in (2), the posterior distribution of θ_i is normal with expectation θ_i^* given by

$$\theta_i^* = \left(\frac{1}{\sigma_i^2} X_i'X_i + \Sigma^{-1}\right)^{-1} \left(\frac{1}{\sigma_i^2} X_i'X_i\hat{\theta}_i + \Sigma^{-1}\mu\right) \tag{8}$$

and covariance matrix $V(\theta_i^*)$ given by

$$V(\theta_i^*) = \left(\frac{1}{\sigma_i^2} X_i'X_i + \Sigma^{-1}\right)^{-1}. \tag{9}$$

For a derivation of this posterior distribution, we refer to Lee and Griffiths (1978). The posterior distribution involves the parameters μ, σ_i^2,and Σ. It appears that there is some sort of recursive problem regarding estimation: to estimate θ_i, we have to know μ, σ_i, and Σ, and to estimate μ, σ_i, and Σ, we have to know θ_i. To overcome this problem, Hu and Maddala (1994) suggest using an empirical

Bayesian approach by estimating the following equations iteratively using $\hat{\theta}_i$ as initial estimation of θ_i^* [1]:

$$\mu^* = \frac{1}{N}\sum_{i=1}^{N}\theta_i^*, \tag{11}$$

$$\sigma_i^{2*} = \frac{1}{T_i - k}(y_i - X_i\theta_i^*)'(y_i - X_i\theta_i^*), \tag{12}$$

and

$$\Sigma^* = \frac{1}{N-1}\sum_{i=1}^{N}(\theta_i^* - \mu^*)(\theta_i^* - \mu^*)'. \tag{13}$$

The algorithm then is as follows:
(a) estimate $\hat{\theta}_i$ using OLS.
(b) set θ_i^* to $\hat{\theta}_i$.
(c) estimate μ^*, σ_i^*, and Σ^* conditional on θ_i^* using equation (10), (11), and (12).
(d) estimate θ_i^* conditional on μ^*, σ_i^*, and Σ^* using equation (8).
(e) goto (c).
(f) return θ_i^*, μ^*, σ_i^*, and Σ^*.

2.3 Gibbs sampling

Although the iterative approach from the previous subsection appears to work out well in practice, see Hu and Maddala (1994), it is sensitive to initialization: if $\hat{\theta}_i$ is very inaccurate, the resulting estimates for μ, σ_i, Σ, and θ_i are also inaccurate. Another, fully Bayesian approach that is less sensitive to initialization is so-called Gibbs sampling. Gibbs sampling is a specific type of Bayesian approach, whereby draws are simulated from the posterior distribution to generate a sample of this distribution rather than deriving it analytically. The density function of the posterior distribution of the parameters is partitioned such that the resulting conditional densities are easy to simulate from. Markov chain theory tells us that repeated application of these conditional densities to an arbitrary density will converge to the density function of the posterior distribution. If the number of iterations is sufficiently large, the impact of initialization is negligibly small. For more information about Markov chains and an in-depth explanation of the Gibbs sampler, we refer to Stern (2000) and Casella and George (1992).

[1] To improve convergence with the iterative procedure, Hu and Maddala (1994) advocate to estimate Σ^* as follows:

$$\Sigma^* = \frac{1}{N-1}\left[D + \sum_{i=1}^{N}(\theta_i^* - \mu^*)(\theta_i^* - \mu^*)'\right], \tag{10}$$

where D denotes a diagonal matrix with small positive entries (e.g. 0.0001).

To implement the Gibbs sampler, we first formulate the conditional distributions of θ_i, μ, σ_i, and Σ to simulate draws from. The conditional distribution of θ_i is simulated from the normal distribution with mean $E(\theta_i)$ given by equation (8) and covariance matrix $V(\theta_i)$ given by equation (9). The conditional distribution of μ is simulated from the normal distribution with mean $E(\mu)$ given by

$$E(\mu) = \frac{1}{N}\sum_{i=1}^{N}\theta_i \tag{14}$$

and variance $V(\mu)$ given by

$$V(\mu) = S/N, \tag{15}$$

where

$$S = \sum_{i=1}^{N}[\theta_i - E(\mu)][\theta_i - E(\mu)]'. \tag{16}$$

The conditional distribution of σ_i is simulated from the χ^2-distribution:

$$\frac{(y_i - X_i\theta_i)'(y_i - X_i\theta_i)}{\sigma_i^2} \sim \chi^2(T_i + 4). \tag{17}$$

The conditional distribution of Σ is simulated from the inverted Wishart distribution:

$$\Sigma \sim iW(S), \tag{18}$$

where

$$S = \sum_{i=1}^{N}(\theta_i - \mu)(\theta_i - \mu)'. \tag{19}$$

To simulate from the inverted Wishart distribution $iW(S)$, a Choleski decomposition is applied:

$$LL' = S^{-1}, \tag{20}$$

where L denotes a lower triangular matrix. Based on this decomposition, Σ is simulated as follows:

$$\Sigma = \left(L(z'z)^{-1}L'\right)^{-1}, \tag{21}$$

where z denotes a $k+1$ by 1 vector of random drawings from a standard normal distribution.

Draws from these distributions are simulated iteratively, conditional on draws of the previous iteration using $\hat{\theta}_i$ as prior for θ_i. The first set of draws are dropped in order to reduce the impact of initial values. The Gibbs sampling algorithm can be summarized as follows:

(a) estimate $\hat{\theta}_i$ using OLS.
(b) set θ_i to $\hat{\theta}_i$.
(c) simulate a draw of μ.
(d) simulate a draw of σ_i conditional on θ_i.
(e) simulate a draw of Σ conditional on θ_i and μ.
(f) simulate a draw of θ_i conditional on μ, σ_i and Σ.
(g) store all draws.
(h) goto (c).
(i) estimate θ_i^*, μ^*, σ_i^*, and Σ^* from stored draws.
(j) return θ_i^*, μ^*, σ_i^*, and Σ^*.

2.4 Summary

If we consider a data set with cross-sectional units characterized by "similar" parameter values, the individual OLS time series estimates as the overall pooled estimate contain information about the unknown parameter values. Shrinkage estimation is a method to incorporate information both from both the time series and cross-section dimensions. Basically, the resulting estimate is the weighted average of the time series estimate and the pooled estimate. We discussed several ways to determine these weights. First, we discussed the Stein-rule estimate whereby the weights depend on the degree of homogeneity of the data. In this approach, the resulting weights are the same for all individual estimates. Second, we discussed a more flexible, empirical Bayesian approach whereby the weights are variable for individual estimates. However, this approach involves some unknown hyperparameters that are estimated in a way that might be sensitive to initialization. In addition, we discussed a fully Bayesian approach whereby simulation is used to estimate the hyperparameters in a way that is robust to initialization effects. However, this approach is more complex and computational intensive. In the next section, we analyze the potential benefits of the three shrinkage approaches for estimating mutual fund performance, by means of a Monte Carlo study.

3 Data generating process

In this section we formulate the data generating process for a Monte Carlo study that uses simulated mutual fund data to compare the shrinkage estimators described in the previous section to standard OLS estimates. The Monte Carlo approach has several attractive properties. First, the true parameter values are known to determine the accuracy of the OLS and the shrinkage estimates. Second, we have complete certainty about the appropriate performance model, so that the results we find can fully be attributed to the estimation method. Third, it allows us to manipulate the underlying data generating process so that we can check the robustness of the employed estimating techniques using different types of data sets.

As mentioned above, several alternative models are used to estimate mutual fund performance. For our simulation exercise, we can restrict attention to a one-factor model. While it is convenient to think of this single factor as the market excess return, corresponding to the CAPM, it may denote any other factor that is considered appropriate. We formulate the one-factor model as

$$r_{it} = \alpha_i + r_{Mt}\beta_i + \varepsilon_{it}, \tag{22}$$

where r_{it} denotes the excess return for fund i a time t, r_{Mt} denotes the excess returns of the market portfolio at time t, α_i denotes the alpha of fund i, and β_i denotes the sensitivity of fund i to the market portfolio. The number of funds in the cross-section equals N, and the number of time periods equals T.

Regarding the error term ε_i, we follow the set-up advocated by Brown et al. (1992) and Horst et al. (2001). Hereby, the idiosyncratic error term is independent of the expected excess return and is assumed to be normally distributed with mean zero and variance σ_i^2, given by

$$\sigma_i^2 = \omega(1 - \beta_i)^2, \tag{23}$$

where ω is a rough approximation to the relationship between non-systematic risk and β_i that is often observed in empirical research.

Returns of the market portfolio are assumed to be normal distributed with mean $E(r_M)$ and variance $V(r_{M,t})$. Usually $E(r_M)$ is assumed to be constant over time, whereas $V(r_{M,t})$ is assumed to be stochastic and to follow some sort of autoregressive process. We employed a specific case of an autoregressive conditional heteroscedasticity (ARCH) model to simulate returns of the market portfolio:

$$V(r_{M,t}) = (1 - \lambda)[r_{M,t-1} - E(r_M)]^2 + \lambda V(r_{M,t-1}), \tag{24}$$

where λ denotes the degree of auto-regression between subsequent variances. For the returns of the market portfolio $T + 10$ observations are generated. The first ten observations are then dropped in order to reduce dependency on initial values.

Due to market equilibrium it can be expected that mutual funds have some similarity between them: as most mutual funds follow the market index, betas are expected to be clustered around one. Alphas are expected to be clustered around zero due to market efficiency. To take this effect into account, mutual fund parameters are assumed to come from a joint normal distribution. To simulate parameters from this multivariate normal distribution, a Choleski decomposition is applied:

$$LL' = \Sigma, \tag{25}$$

where Σ denotes the 2 by 2 covariance matrix, and L denotes a lower triangular matrix. Based on this decomposition, the parameters α_i and β_i are simulated as follows:

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \mu + Lz, \tag{26}$$

where μ denotes the 2 by 1 vector of means of the parameters, and z denotes a 2 by 1 vector of random drawings from a standard normal distribution.

The employed constant values are set in such a way that the model mimics typical mutual fund data characteristics. Mutual fund data are often characterized by short time series. We choose to simulate excess returns of mutual funds for a period of one year on monthly basis. The number of time periods is set to $T = 12$. Another characteristic of mutual fund data is the large cross-section. The number of mutual fund in the cross-section is set to $N = 250$. Regarding the error term, we follow Horst et al. (2001) and set ω equal to 0.0065. A realistic expected value of the excess return of the market portfolio is about 8 percent per year. Monthly expected returns are set to $E(r_M) = 0.0065$. A realistic value of market volatility is about 20 percent per year. The variance of monthly expected returns is set to $V(r_{M,t}) = 0.0035$. The degree of auto-regression between subsequent variances is set to $\lambda = 0.6$. Alphas are typically very small with an expected value of zero and a standard deviation of 1 percent per year. Betas are clustered around one with a standard deviation of 0.1. The expected value and the variance of alpha and beta are set to

$$\mu = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{27}$$

and

$$\Sigma = \begin{pmatrix} 0.0030 & 0.0000 \\ 0.0000 & 0.1000 \end{pmatrix}^2. \tag{28}$$

Now, we implement this data generating process and the shrinkage estimators we described in the previous section in the matrix programming language Ox (Doornik, 1998). The program first generates the parameters of the mutual funds. Second, a time series of excess returns are generated for the market index and the funds. Based on these returns, the parameters are estimated using OLS and the shrinkage estimators. We run this simulation 10,000 times [2] and analyze the

[2] The Monte Carlo experiment is run on a Pentium-3 500 megahertz computer.

accuracy and the robustness of the estimates as well as the implications of their use in persistence analysis. We set the iterative empirical Bayes estimator to iterate 10 times, and the Gibbs sampler to iterate 50 times. The first 10 draws from the Gibbs sampler are dropped. These numbers of iterations might seem surprisingly small, but it appears that the initial OLS estimate is a reasonable good starting value in general, so that the convergence rate of the estimates if quite large. The results of this Monte Carlo experiment are presented and discussed in the following section.

4 Results

4.1 Accuracy

The first result we are interested in is the accuracy of the OLS and shrunk estimates. One way to evaluate the accuracy of an estimator is to consider the estimator's root mean square error (RMSE). RMSE is the root of the average squared distance between $\hat{\theta}$ and the true parameter value θ. In fact, RMSE is a measure of the loss from using $\hat{\theta}$ as an estimator of θ so that the mathematical average of this loss yield the average loss or risk of using $\hat{\theta}$ to estimate θ. We calculate the RMSE for the whole cross-section:

$$RMSE_{\theta_i} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\alpha_i - \hat{\alpha_i})^2}. \tag{29}$$

In this way, RMSE measures the loss from using $\hat{\theta}$ as an estimator of θ to describe the whole data set. Secondly, we are interested in the computational complexity of the employed shrinkage estimators. Therefore, we consider the lapsed time for all estimators per simulation. The lapsed time is measured in seconds. For all simulations, we calculate RMSE and the lapsed time of the OLS and the shrunk estimates. In Table 1 we present the average RMSE per simulation for all estimators, the difference between the OLS and the shrunk estimates, and the average lapsed time for all estimators. From these results, we can infer the following: the average $RMSE$ of the OLS estimate is 0.2919 percent per month. That is about 1 percent per year. The average $RMSE$ of the Stein-rule, iterative empirical Bayes, and Gibbs sampler estimates are 0.2671 percent, 0.1713 percent, and 0.1703 percent per month respectively. That is about 0.9 percent for the Stein-rule estimates and 0.6 percent for the iterative empirical Bayes and Gibbs sampler estimates per year respectively. It seems that OLS estimates are highly inaccurate: the value of $RMSE$ is relatively large as compared to alpha, which has an average of zero and a standard deviation of 1 percent per year. Estimates can therefore easily be several times smaller or larger than the true values. Furthermore, because alpha generally has a small value, it is not possible to determine the sign of alpha in some cases. All shrunk estimates are more accurate, especially the iterative empirical Bayes and the Gibbs sampler estimates. As compared to OLS estimates, Stein-rule estimates are about 10 percent and iterative empirical Bayes and Gibbs sampler estimates are about 40 percent more accurate. Although the iterative empirical Bayes and the Gibbs sampler estimates are about as accurate, it seems that the Gibbs sampler is somewhat more stable as the difference with the OLS estimate yields less variance. Regarding the average lapsed time per estimate, it seems that the time necessary is a trivial matter as all estimates require less than a second to compute.

Method	RMSE avg.	difference avg.	std.dev.	min.	max.	lapsed time (s) avg.
normal						
OLS	0.2919					0.00
Stein	0.2671	0.0248*	0.0126	0.0000	0.0807	0.12
Bayes	0.1713	0.1205*	0.0921	0.0297	0.5509	0.19
Gibbs	0.1703	0.1216*	0.0871	0.0385	0.5404	0.66

Table 1: average $RMSE$ of OLS and shrunk estimates, statistics of difference between both approaches, and average lapsed time per estimator. Notes: $*$ denotes significance at a 5 percent level. Significance is measured using a Wilcoxon rank sum test

4.2 Robustness

In the previous subsection we found that shrinkage estimators are substantially more accurate than OLS estimates in our simulated setting. In this sections, we evaluate this accuracy under a range of different settings. The reason that shrunk estimates are more accurate in our simulated setting is because they incorporate cross-sectional similarity by, for example, assuming that cross-sectional parameter values are normally distributed. Because the simulated data is generated in such a way that cross-sectional parameter values are indeed normally distributed, it is to be questioned whether these shrinkage estimators are still more accurate when parameter values are non-normally distributed. To evaluate the sensitivity of the employed estimators to non-normal values of alpha and beta, we run an additional simulation where alpha and beta are generated from the Student's t- and the lognormal distribution. Again, we calculate the average RMSE, and the difference between the OLS and the shrunk estimates, but now under different distributions of the parameters. The results are listed in Table 2. By simulating extreme- and asymmetric distributed parameter values, two effects are triggered: first, because of the feedback between β_i and the error term, $RMSE$ increases for all estimators as beta can now have extreme values. Second, because the iterative empirical Bayes estimator and the Gibbs sampler assume that alpha and beta are normally distributed, $RMSE$ might increase if these estimators are not robust to these underlying assumptions. As the OLS estimator makes no assumptions on the properties of the distribution of alpha and beta, OLS estimates are only affected by the feedback between β_i and the error term. Therefore, we can use the difference in $RMSE$ between shrunk- and OLS estimates as a benchmark for robustness under different distributions of alpha and beta. The results indicate that this difference increases under all simulated distributions. In fact, it seems that the more extreme- and asymmetric distributed values are generated, the more this difference increases. Apparently, the first effect strongly overcompensates the second effect. So even if the parameters are non-normally distributed, shrinkage estimators are significantly more accurate than standard OLS estimates.

Method	RMSE	difference			
	avg.	avg.	std.dev.	min.	max.
t-dist, 5df					
OLS	0.3837				
Stein	0.3519	0.0318*	0.0233	0.0000	0.1663
Bayes	0.2137	0.1701*	0.1937	-0.3001	1.2503
Gibbs	0.2084	0.1753*	0.1779	-0.2673	1.1491
t-dist, 3df					
OLS	0.5346				
Stein	0.4862	0.0483*	0.0469	0.0000	0.3938
Bayes	0.2833	0.2513*	0.4564	-0.1080	3.8208
Gibbs	0.2509	0.2837*	0.4275	-0.0678	3.8458
lognormal					
OLS	0.7872				
Stein	0.7020	0.0851*	0.0518	0.0000	0.3015
Bayes	0.3987	0.3885*	0.3427	-0.3670	1.7932
Gibbs	0.3395	0.4477*	0.3328	-0.2850	1.8735

Table 2: average $RMSE$ of OLS and shrunk estimates and statistics of difference between both approaches under different distributions of the parameters. Notes: $*$ denotes significance at a 5 percent level. Significance is measured using a Wilcoxon rank sum test

Another factor that influences the accuracy of the shrinkage estimators is the size of the cross-section: the larger the size of the cross-section, the more information the pooled estimate contains. Therefore, it can be questioned whether shrinkage estimators still remain more accurate than OLS estimates if the size of the cross-section decreases. We run simulations with different sizes of the cross-section to evaluate the sensitivity of the employed estimators to this factor. The results are listed in Table 3. It appears that, even if the size of the cross-section decreases, the shrinkage estimators remain significant more accurate.

A similar factor that influences the accuracy of the shrinkage estimators is the length of the time series. It can be questioned whether shrinkage estimators

Method	RMSE avg.	difference avg.	std.dev.	min.	max.	lapsed time avg.
N = 50						
OLS	0.2578					0.00
Stein	0.2386	0.0193*	0.0074	0.0038	0.0282	0.01
Bayes	0.1680	0.0898*	0.0424	0.0254	0.1644	0.04
Gibbs	0.1775	0.0803*	0.0324	0.0360	0.1283	0.10
N = 100						
OLS	0.2595					0.00
Stein	0.2411	0.0185*	0.0078	0.0010	0.0302	0.02
Bayes	0.1703	0.0893*	0.0383	0.0082	0.1308	0.07
Gibbs	0.1697	0.0898*	0.0282	0.0329	0.1277	0.24
N = 500						
OLS	0.3008					0.01
Stein	0.2740	0.0268*	0.0089	0.0159	0.0457	0.58
Bayes	0.1710	0.1299*	0.0616	0.0618	0.2742	0.38
Gibbs	0.1687	0.1322*	0.0579	0.0719	0.2768	1.75

Table 3: average $RMSE$ of OLS and shrunk estimates, statistics of difference between both approaches under different sizes of the cross-section, and average lapsed time per estimator. Notes: $*$ denotes significance at a 5 percent level. Significance is measured using a Wilcoxon rank sum test

still remain more accurate than OLS estimates if the length of the time series increases. We repeat the experiment, but now with different lengths of the time series. The results are listed in Table 4. It seems that the shrunk estimates remain significantly more accurate, even if the length of the time series is large.

We also examined the sensitivity of the shrinkage estimators to the size of the error term and of the variance of alpha. The results are listed in Table 5 and 6. Again, it seems that the shrinkage estimators remain significantly more accurate, especially when the error term is relatively large or the variance of alpha is relatively small.

4.3 Persistence Analysis

We can now conclude that shrinkage estimators are substantially more accurate than standard OLS estimates under a wide range of settings. In this section we investigate what the implications are from using shrinkage estimation as an alternative approach to persistence analysis. The typical approach to analyze persistence in mutual fund performance is to estimate alpha for each fund for a given time interval. All funds are ranked into deciles based on their estimation of alpha. We call this the ranking period. Next, the Sharpe ratio is estimated for each fund for the subsequent time period. We call this the post-ranking period.

Method	RMSE avg.	difference avg.	std.dev.	min.	max.	lapsed time
T = 24						
OLS	0.2159					0.01
Stein	0.2106	0.0054*	0.0024	0.0020	0.0104	0.29
Bayes	0.1383	0.0776*	0.0573	0.0152	0.2117	0.19
Gibbs	0.1390	0.0769*	0.0537	0.0200	0.2066	0.90
T = 36						
OLS	0.1931					0.01
Stein	0.1904	0.0026*	0.0016	0.0002	0.0064	0.47
Bayes	0.1182	0.0749*	0.0778	0.0050	0.2941	0.20
Gibbs	0.1215	0.0715*	0.0719	0.0126	0.2744	1.10
T = 60						
OLS	0.1496					0.01
Stein	0.1488	0.0008*	0.0003	0.0002	0.0015	0.93
Bayes	0.0973	0.0523*	0.0435	0.0008	0.1610	0.25
Gibbs	0.0994	0.0502*	0.0415	0.0006	0.1521	1.66

Table 4: average $RMSE$ of OLS and shrunk estimates, statistics of difference between both approaches under different lengths of the time series, and average lapsed time per estimator. Notes: $*$ denotes significance at a 5 percent level. Significance is measured using a Wilcoxon rank sum test

The Sharpe ratio is defined as

$$Sharpe_i = \frac{E(r_i)}{s(r_i)}, \tag{30}$$

where $E(r_i)$ denotes the mean of r_i over T time periods, and $s(r_i)$ denotes the standard deviation of r_i over T time periods. This approach may answer the question, for example, to what extent the top decile can be expected to perform above average in the future. Crucial to this type of analysis is the ranking ability of the employed estimator.

We evaluate this ability under full persistence by following the setup of Busse and Irvine (2002). In order to examine whether the relation between the estimation of alpha during the ranking period and subsequent Sharpe-ratio is statistically significant, we compute Spearman rank correlation coefficients between the ranking period decile ranking and the post-ranking period Sharp-ratio decile ranking per simulation. The Spearman rank correlation coefficient R^2 is defined as

$$R^2 = 1 - \frac{6\sum d^2}{n(n^2-1)}, \tag{31}$$

where $\sum d^2$ is the sum of the squared differences between ranks of the ranking- and the post-ranking period, and n is the number of ranks. In our simulation,

Method	RMSE	difference			
	avg.	avg.	std.dev.	min.	max.
ω = 0.005					
OLS	0.2371				
Stein	0.2203	0.0169*	0.0060	0.0039	0.0259
Bayes	0.1566	0.0805*	0.0299	0.0315	0.1288
Gibbs	0.1568	0.0804*	0.0263	0.0331	0.1184
ω = 0.01					
OLS	0.3353				
Stein	0.3003	0.0351*	0.0117	0.0102	0.0507
Bayes	0.1929	0.1424*	0.0436	0.0651	0.2061
Gibbs	0.1874	0.1479*	0.0391	0.0771	0.2089

Table 5: average $RMSE$ of OLS and shrunk estimates and statistics of difference between both approaches under different sizes of the error term. Notes: $*$ denotes significance at a 5 percent level. Significance is measured using a Wilcoxon rank sum test

the number of time periods of the ranking period is $T_{pre} = 12$, and the number of time periods of the post-ranking period is $T_{post} = 12$. For all simulations, we calculate the Spearman rank correlation coefficient per estimator. In Table 7 we present the average Spearman rank correlation coefficient per simulation for all estimators, and the difference between the OLS and the shrunk estimates. As the Stein-rule applies a constant degree of shrinkage on all time series estimates, resulting rankings are equivalent to rankings based on standard OLS. Persistence studies based on the Stein-rule are therefore equivalent to those based on OLS. However, persistence studies based on iterative empirical Bayes and Gibbs sampling appear to be significantly more reliable. To examine the economic relation between the estimation of alpha during the ranking period and subsequent Sharpe-ratio, the difference in post-ranking period Sharp ratio between top and bottom decile is computed. For all simulations, we calculate the difference in Sharp ratio between top and bottom per estimator. In Table 8 we present the average Sharpe ratio difference per simulation for all estimators, and the difference between the OLS and the shrunk estimates. These results are consistent with the previous experiment.

5 Conclusions

The estimation of short-run performance measures of mutual funds is characterized by large standard errors, implying that corresponding mutual fund rankings are to a large extent based on luck rather than skills. In general, the use of shrinkage estimation techniques reduces the impact of luck components upon performance measures and accordingly improves the cross-section of estimated

Method	RMSE avg.	difference avg.	std.dev.	min.	max.
σ_α = *0.005*					
OLS	0.2704				
Stein	0.2579	0.0125*	0.0041	0.0048	0.0194
Bayes	0.2016	0.0688*	0.0317	0.0175	0.1135
Gibbs	0.2032	0.0672*	0.0302	0.0157	0.1122
σ_α = *0.001*					
OLS	0.2704				
Stein	0.2327	0.0376*	0.0141	0.0062	0.0566
Bayes	0.0801	0.1903*	0.0330	0.1285	0.2459
Gibbs	0.0992	0.1711*	0.0314	0.1130	0.2253

Table 6: average $RMSE$ of OLS and shrunk estimates and statistics of difference between both approaches under different sizes of variance of alpha. Notes: $*$ denotes significance at a 5 percent level. Significance is measured using a Wilcoxon rank sum test

Method	Spearman avg.	difference avg.	std.dev.	min.	max.
OLS	0.54				
Stein	0.54	0.00*	0.00	0.00	0.00
Bayes	0.61	0.07*	0.05	-0.05	0.25
Gibbs	0.60	0.06*	0.04	-0.01	0.23

Table 7: average Correlation coefficient of OLS and shrunk estimates, and statistics of difference between both approaches. Notes: $*$ denotes significance at a 5 percent level. Significance is measured using a Wilcoxon rank sum test

fund alphas. In this paper we evaluated three alternative shrinkage estimators and compared them with traditional OLS estimates for performance. The Stein rule estimator takes a weighted average of the individual OLS estimates and an overall pooled estimate, where the weights are identical across funds and based upon the degree of similarity of the individual estimates. The two Bayesian approaches allow the weights to differ across funds depending upon the "likelihood" of the estimated individual alpha. Typically, more extreme estimates for performance will be "shrunk" more towards the overall mean.

While the Stein rule estimator does not affect the relative position of funds in their cross-section, the Bayesian approaches may lead to a different ranking of mutual fund performance. Based on a Monte Carlo study, we analyze the mean-squared errors of the different alpha estimates, and the ability to reproduce the original performance ranking that was used to generate the data.

The Monte Carlo study leads to the following conclusions. If we estimate alphas using standard OLS, the average RMSE is about 1 percent per year in our basic set up. These estimates are highly inaccurate as the true alphas have

Method	Sharpe avg.	difference avg.	std.dev.	min.	max.
OLS	0.8127				
Stein	0.8127	0.0000*	0.0000	0.0000	0.0000
Bayes	1.3839	0.5712*	0.7146	-0.0679	3.2706
Gibbs	1.2261	0.4134*	0.5115	-0.0432	2.3274

Table 8: average difference in Sharp ratio between top and bottom of OLS and shrunk estimates, and statistics of difference between both approaches. Notes: $*$ denotes significance at a 5 percent level. Significance is measured using a Wilcoxon rank sum test

an average of zero and a standard deviation of 1 percent per year. As a result, estimates can easily be several times smaller or larger than the actual values and the signs may differ in a substantial number of cases. The employed shrinkage estimators are all significantly more accurate with an average RMSE of 0.9 percent per year for the Stein-rule estimator and only 0.6 percent for iterative empirical Bayes and the Gibbs sampler. When we evaluate the sensitivity of the employed shrinkage estimators to extreme- and asymmetric distributed values of alpha and beta, we find that the superior performance of the shrinkage estimators is robust to parameters that come from the Student's t- and the lognormal distribution.

We also performed persistence analysis using the different approaches. As the Stein-rule applies a constant degree of shrinkage on all time series estimates, resulting rankings are equivalent to rankings based on standard OLS. However, persistence studies based on iterative empirical Bayes and Gibbs sampling appear to be significantly more reliable: we evaluate the statistical implications by estimating rank correlations between the estimated alphas and the subsequent Sharpe ratios. Rank correlations based on iterative empirical Bayes and Gibbs sampling are significantly more accurate. Further, we evaluated the economic implications by calculating the average difference between the subsequent Sharpe ratio of the top- and bottom ranking decile. We find that this difference is significant larger for rankings based on empirical Bayes and Gibbs sampling.

In general, the results in this paper indicate that the use of shrinkage estimators, exploiting the entire cross-section of mutual fund returns, is a very fruitful approach. Performance estimates are substantially more accurate, while performance rankings improve compared to traditional OLS measures.

References

K. Baks, A. Metrick, and J. Wachter. Bayesian performance evaluation. *working paper*, 1999.

R. Blattberg and E. George. Shrinkage estimation of price and promotional elasticities: Seemingly unrelated equations. *The Journal of the American Statistical Association*, 86:304–315, 1991.

M. E. Blume. On the assesment of risk. *The Journal of Finance*, 26:1–10, 1971.

S. J. Brown, W. Goetzmann, R. G. Ibbotson, and S. A. Ross. Survivorship bias in performance studies. *The Review of Financial Studies*, 5:553–580, 1992.

J. Busse and P. Irvine. Bayesian alphas and mutual fund persistence. *working paper*, 2002.

M. Carhart. On persistence in mutual fund performance. *The Journal of Finance*, 52:57–82, 1997.

G. Casella and E. George. Explaining the gibbs sampler. *The American Statistician*, 46:167–174, 1992.

J. A. Doornik. *Object-Oriented Programming Using Ox*. Timerlake Consultants, 1998.

E. F. Fama and K. R. French. The cross-section of expected stock returns. *The Journal of Finance*, 47:427–465, 1992.

D. Hendricks, J. Patel, and R. Zeckhauser. Hot hands in mutual funds: Short-run persistence of relative performance, 1974-1988. *The Journal of Finance*, 48:93–130, 1993.

J. R. T. Horst, T. E. Nijman, and M. Verbeek. Eliminating look-ahead bias in evaluating persistence in mutual fund performance. *The Journal of Empirical Finance*, 8:345–373, 2001.

W. Hu and G. Maddala. Estimation and prediction problems in dynamic heterogeneous panel data models. *working paper, Ohio State University, Dept. of Economics*, 1994.

L. Lee and W. Griffiths. The prior likelihood and best linear unbiased prediction in stochastic coefficient linear models. *working paper, University of Minnesota, Dept. of Economics*, 1978.

G. Maddala, R. Trost, H. Li, and F. Joutz. Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators. *The Journal of Business and Economic Statistics*, 15, 1997.

B. Malkiel. Returns from investing in equity fund 1971 to 1991. *The Journal of Finance*, 50:549–572, 1995.

L. Pastor and R. Stambaugh. Mutual fund performance and seemingly unrelated assets. *Journal of Financial Economics, forthcoming*, 2001.

S. Stern. *Simulation-Based Inference in Econometrics.* The University of Cambridge Press, 2000.

O. A. Vasicek. A note on using cross-sectional information in bayesian estimation of security betas. *The Journal of Finance*, 28:1233–1239, 1973.