JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS VOL. 21, NO. 3, SEPTEMBER 1986

Bayes-Stein Estimation for Portfolio Analysis

Philippe Jorion*

Abstract

In portfolio analysis, uncertainty about parameter values leads to suboptimal portfolio choices. The resulting loss in the investor's utility is a function of the particular estimator chosen for expected returns. So, this is a problem of simultaneous estimation of normal means under a well-specified loss function. In this situation, as Stein has shown, the classical sample mean is inadmissible. This paper presents a simple empirical Bayes estimator that should outperform the sample mean in the context of a portfolio. Simulation analysis shows that these Bayes-Stein estimators provide significant gains in portfolio selection problems.

I. Introduction

In medio virtus

One of the fundamental propositions of the modern virtus theory of finance is that security risk has to be considered in the context of a portfolio. It is astonishing then that estimation techniques in finance have not recognized the implications of this result for efficient estimation of unknown parameters. In the context of a portfolio, using sample means to estimate expected returns amounts to ignoring information contained in other series, and could be compared to assessing the risk of a security by looking at the variance of its return, rather than at its contribution to overall portfolio risk.

This paper presents an application of shrinkage estimation to portfolio selection problems. Shrinkage estimators have already been used in finance (see [31], [7], and [19]), but always on an ad hoc basis. This paper provides a sound rationale for such estimators, and illustrates the extent of possible gains over classical estimators.

The application of portfolio analysis à la Markowitz [25] traditionally proceeds in two steps. First, the moments of the distribution of returns are estimated from a time-series of historical returns; then the mean variance portfolio selection problem is solved separately, as if the estimates were the true parameters. This

* Graduate School of Business, Columbia University, New York, NY 10027. This paper draws on part of the author's dissertation at the University of Chicago, and was presented at the 1983 NBER-NSF Conference on Bayesian Inference in Econometrics. The author is grateful to Arnold Zellner for useful discussions and comments and an anonymous *JFQA* referee who also helped to improve the paper. The research was supported by the Collège Interuniversitaire d'Etudes Doctorales dans les Sciences du Management.

"certainty equivalence" viewpoint, in which the sample estimates are treated as the true values, has been criticized by Barry[1], Brown [11], and Klein and Bawa [22], who advocate a Bayesian approach that explicitly incorporates estimation error. But their conclusions should be carried further, which is what the present study proposes to do.

The impact of parameter uncertainty on optimal portfolio selection has been recognized by a number of authors (see [17], [13], and [20]) who show that the practical application of portfolio analysis is seriously hampered by estimation error, especially in expected returns. Variances and covariances are also unknown, but are more stable over time, as pointed out by Merton [27].

In this context, the relevant measure of estimation risk is the utility loss due to a portfolio choice based on sample estimates, rather than on true values. This loss is a function of the estimator chosen for the population moments. Consequently, one should select an estimator with average minimizing properties *relative to this loss function*. Brown [11] provided a Bayesian correction based on a diffuse prior, which reduces estimation risk, but the estimator in this case is still the sample mean, which too often takes extreme values. Further, choosing sample means does not fully exploit the multivariate nature of the problem. The issue here is not to estimate each expected return separately, in which case the sample mean would be optimal, but rather to minimize the impact of estimation risk on optimal portfolio choice. Thus, the portfolio context should be central to the estimation procedure.

Instead of the sample mean, an estimator obtained by "shrinking" the means toward a common value is proposed, which should lead to decreased estimation error with more than two assets in the portfolio. This result can be traced to the inadmissibility of the sample mean, which was first proved by Stein [29] and extended by Brown [8]. It stems from the fact that the effect of estimation error for all assets is summarized into one loss function, which should be minimized as a whole rather than each component separately.

Section II develops the topic of estimation risk, and Section III reviews the original Stein estimator and its extensions. An empirical Bayes interpretation is presented in Section IV. The shrinkage effect is explained by an informative prior. The parameters of the prior are not prespecified but, rather, are computed from the data themselves.

Section V illustrates the gains from Bayes-Stein estimation. By simulation analysis, it is shown that this estimator drastically reduces estimation error: expressed in risk-free equivalent return, the gain over the Bayes diffuse prior is of the order of a few percent per annum for sample sizes below 50. Some concluding remarks are offered in Section VI.

II. Estimation Risk

Traditionally, statistical estimation has been kept separate from portfolio decisions, mainly because portfolio choice has been analyzed in the "certainty equivalence" framework, in which the underlying moments are assumed known. But this two-step procedure is not optimal from an estimation viewpoint: efficiency can be improved by directly considering the effect of parameter uncertainty on the investor's utility.

First, estimation risk should be defined. In a one-period model, the usual rationality axioms lead the investor to maximize the expected utility of his or her end-of-period wealth. In terms of rates of return, the control problem is to choose a set of weights $\underline{q}$ so as to maximize the expected utility of the return on the portfolio $z = \underline{q}'\underline{r}$, where $\underline{r}$ is the vector of future observations,

$$EU(z) = \int U(z)p(z \mid \underline{\theta})dz \,, \tag{1}$$

subject to a feasibility constraint.

The problem has two components: (i) a utility function $U(z)$, generally characterized by a functional form and a set of parameters, both of which can be different across investors; and (ii) the conditional distribution of rates of return, conditioned on a set of parameters $\underline{\theta}$, unknown for all practical purposes.

In the *certainty equivalence* framework, one assumes that $\underline{\theta}$ equals its estimate $\hat{\underline{\theta}}(\underline{y})$, based on some estimator, defined as a function of the observations $\underline{y}$,

$$\max_{\underline{q}} E_{\underline{y}}\left[U(z) \mid \underline{\theta} = \hat{\underline{\theta}}(\underline{y})\right] \,. \tag{2}$$

This approach obviously ignores the issue of estimation risk, or parameter uncertainty.

The Bayesian solution to this problem, as first suggested by Zellner and Chetty [33], is to define optimal portfolio choice in terms of the predictive density function. The latter is obtained after integrating out the unknown parameter $\underline{\theta}$, which explicitly takes into account uncertainty about $\underline{\theta}$. The investor's problem can be described as the maximization of the *unconditional* expected utility of his portfolio

$$\max_{\underline{q}} E_{\underline{\theta}}\left[E_{\underline{y} \mid \theta}[U(z) \mid \underline{\theta}]\right] \,, \text{ with}$$

$$E_{\underline{\theta}}\left[E_{\underline{y} \mid \hat{\underline{\theta}}}[U(z) \mid \underline{\theta}]\right] = \int_{\underline{\theta}}\int_{\underline{y}} U(z)p(z \mid \underline{\theta})dz\, p\left(\underline{\theta} \mid \underline{y}, I_0\right)d\underline{\theta}$$

$$= \int_{\underline{y}} U(z)\left[\int_{\underline{\theta}} p(z \mid \underline{\theta})p\left(\underline{\theta} \mid \underline{y}, I_0\right)d\underline{\theta}\right]dz \,.$$

The term between brackets is defined as the *predictive density function* of $z = \underline{q}'\underline{r}$

$$p(z \mid \underline{y}) = \int_{\underline{\theta}} p(z \mid \underline{\theta})p\left(\underline{\theta} \mid \underline{y}, I_0\right)d\underline{\theta} \,, \tag{3}$$

where $p(\underline{\theta} \mid \underline{y}, I_0)$ is the posterior density function of $\underline{\theta}$, given the data and the prior information I_0,

$$p\left(\underline{\theta} \mid \underline{y}, I_0\right) \div f(\underline{y} \mid \underline{\theta})p\left(\underline{\theta} \mid I_0\right) \,.$$

As Klein and Bawa [22] have shown, this approach is optimal according to the expected utility von Neumann-Morgenstern axioms.

In a mean-variance framework the objective function can be reduced to a derived utility function

$$(4) \qquad EU(z) = F\left(\mu_z, \sigma_z^2\right),$$

where $\mu_z = \underline{q}'\underline{\mu}$ and $\sigma_z^2 = \underline{q}' \, \Sigma \, \underline{q}$ are the expected return and variance of the portfolio. The control problem is to choose the vector of investment proportions $\underline{q}$ so as to maximize expected utility, subject to the constraint that the weights must sum to one, for instance.

If the distribution moments $\underline{\theta} = (\underline{\mu}, \Sigma)$ are known, the choice must be optimal

$$(5) \; F\left(\mu_z^*, \sigma_z, 2^*\right) = F(\underline{q}^*(\underline{\theta}) \mid \underline{\mu}, \Sigma) \equiv F(\underline{q}^{*\prime}\underline{\mu}, \underline{q}^{*\prime}\Sigma\underline{q}^*) = F_{\text{MAX}} \, .$$

On the other hand, if the parameters $\underline{\theta}$ are unknown, the portfolio choice $\underline{q}$ will be made on the basis of the sample estimate $\hat{\underline{\theta}}(\underline{y})$. The expected utility, measured in terms of the *true* underlying distribution, will necessarily be lower than before

$$F\left(\hat{\underline{q}}\left(\hat{\underline{\theta}}(\underline{y})\right) \mid \underline{\mu}, \Sigma\right) \equiv F(\hat{\underline{q}}'\underline{\mu}, \hat{\underline{q}}'\Sigma\hat{\underline{q}}) \leqslant F_{\text{MAX}} \, .$$

Clearly, this loss in utility is due to parameter uncertainty. Following Brown [11], the loss due to estimation risk can be measured as

$$(6) \qquad L(\underline{q}^*, \hat{\underline{q}}) = \frac{F_{\text{MAX}} - F(\hat{\underline{q}})}{|F_{\text{MAX}}|} \, .$$

Figure 1 illustrates the utility loss due to estimation risk. If the investor knew the true parameters, he would choose the portfolio represented by point A, where the weights q^* are optimal relative to the true frontier (the solid line). Unfortunately, he only observes sample estimates, and selects portfolio B, with composition $\hat{q}$, which is optimal relative to the estimated frontier (the dashed line). However, this choice $\hat{q}$ is suboptimal relative to the *true* parameters: for point C, $F(\hat{\underline{q}}) \leqslant F_{\text{MAX}}$. The difference between these values can be attributed to estimation risk; it can also be expressed in risk-free equivalent return, by transforming the level of expected utility F into a risk-free rate R

$$(7) \qquad F\left(\mu_z, \sigma_z^2\right) = F(R, 0) \, .$$

Various estimators $\hat{\underline{\theta}}(\underline{y})$ imply various portfolio choices $\hat{\underline{q}}(\hat{\underline{\theta}}(\underline{y}))$ and, thus, various losses $L(\underline{\theta}, \hat{\underline{\theta}}, (\underline{y}))$. This leads to a well-defined loss function for the estimator $\hat{\underline{\theta}}(\underline{y})$ viewed as a function $t(\cdot)$ of the data:

1. for $\hat{\underline{\theta}}(\underline{y}) = \underline{\theta}$, the loss is zero,
2. for any $\hat{\underline{\theta}}(y) \neq \underline{\theta}$, the loss is nonnegative.

Because the loss is a function of random elements, it cannot be minimized

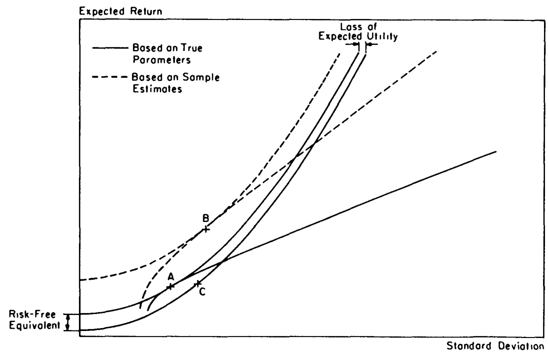

FIGURE 1

Portfolio Choice with Estimation Error

as such. In sampling theory, the *risk function* for an estimator $t(\cdot)$ is defined as the expected loss over repeated samples

(8) $$R_t(\underline{\theta}) = \int L\left(\underline{\theta}, \hat{\underline{\theta}}(\underline{y})\right) f(\underline{y} \mid \underline{\theta}) d\underline{y} .$$

A decision rule $t_0(\cdot)$ is said to be *inadmissible* if there exists another rule $t_1(\cdot)$ with at least equal and sometimes lower risk for any value of the true unknown parameter $\underline{\theta}$.

Thus, a reasonable minimum requirement for any estimator is admissibility. The central thesis of this paper is that the usual sample mean is not admissible for portfolio estimation.

III. Stein Estimation

Consider the problem of estimating $\underline{\mu}$, the vector of means of N normal random variables, distributed as

$$\underline{y}_t \sim NID(\underline{\mu}, \Sigma) , \qquad t = 1, \ldots, T ,$$

where the covariance matrix is assumed known. For N greater than 2, Efron and Morris [16], generalizing Stein's [29] result, showed that the maximum-likelihood estimator $\hat{\underline{\mu}}_{\text{ML}}(\underline{y})$, which is also the vector of sample means $\underline{Y}$, is inadmissible relative to a quadratic loss function

(9) $$L(\underline{\mu}, \hat{\underline{\mu}}(\underline{y})) = (\underline{\mu} - \hat{\underline{\mu}}(\underline{y}))' \Sigma^{-1} (\underline{\mu} - \hat{\underline{\mu}}(\underline{y})) .$$

The use of this loss function is relatively widespread because it leads to tractable results. It is interesting to study because, in the univariate case, the optimal estimator is the sample mean. Also, a quadratic loss is generally a good local approximation of a more general loss function expanded in a Taylor series. For repeated observations, the so-called James-Stein estimator

(10) $$\hat{\underline{\mu}}_{\text{JS}}(\underline{y}) = (1 - \hat{w})\underline{Y} + \hat{w} Y_0 \underline{1} ,$$

where $\hat{w}$ is defined as

(11) $$\hat{w} = \min\left(1, \frac{(N-2)/T}{\left(\underline{Y} - Y_0\underline{1}\right)'\Sigma^{-1}\left(\underline{Y} - Y_0\underline{1}\right)}\right),$$

has *uniformly lower risk* than the sample mean $\underline{Y}$. This estimator is also called a *shrinkage estimator,* since the sample means are multiplied by a coefficient $(1 - \hat{w})$ lower than one. Further, the estimator can be shrunk toward *any point* Y_0, and still have lower risk than the sample mean, but the gains are greater when Y_0 is close to the true value. For negative values of $(1 - \hat{w})$, setting the coefficient equal to zero leads to an improved estimator: this is the positive part rule.[1] Note that this estimator is biased and nonlinear, since the shrinkage factor is itself a function of the data.

The superiority of the James-Stein estimator is a startling result. Indeed, statisticians have been slow to recognize this powerful new statistical technique, in spite of Lindley's [23] early description of it as "one of the most important statistical ideas of the decade."[2]

Basically, the result stems from the *summation* of the components of the loss function: Stein's estimators achieve uniformly lower risk than the maximum likelihood estimator, allowing increased risk to some individual components of the vector $\underline{\mu}$. As a result, the inadmissibility of the sample mean has been extended by Brown ([8], [9], and [10]) to other loss functions under surprisingly weak conditions.

Unfortunately, proof of inadmissibility does not necessarily lead to the construction of better estimators, and the computation of the risk function is an arduous task. Berger [3], however, developed an approach that leads to improved estimators for polynomial loss functions. For a loss function that is the square of the usual quadratic loss, he finds that a shrinkage factor of the form

(12) $$\hat{w} = \frac{b}{d + \left(\underline{Y} - Y_0\underline{1}\right)'T\Sigma^{-1}\left(\underline{Y} - Y_0\underline{1}\right)},$$

with $0 \leq b \leq 2(N-2)$ and some weak conditions on d, leads to an estimator better than the sample mean. Estimators of this form tend to be very robust with respect to the exact functional form of the loss, as Brown [8] demonstrated.

It has been shown that the assumptions of known σ^2 and of normality are not critical, but Berger [5] has indicated that the improvement is most significant in "symmetric" situations, where variances are similar across series. As standard deviations of stock returns are large relative to sample means, and similar across stocks, we expect that Stein estimation should lead to significant improvement over the sample mean.

[1] Berger and Bock [6] discuss methods for improving Stein estimators, based on eliminating singularities.

[2] See [23], p. 285, and also the explanations advanced by Efron and Morris [14] for the resistance to this new concept.

IV. The Empirical Bayes Approach

The surprising results found by Stein have been given an interesting Bayesian interpretation. Consider the following informative conjugate prior for $\underline{\mu}$

$$(13) \qquad p(\underline{\mu} \mid \eta,\lambda) \div \exp\left[-\left(\frac{1}{2}\right)(\underline{\mu} - \underline{1}\eta)'\left(\lambda\Sigma^{-1}\right)(\underline{\mu}-\underline{1}\eta)\right],$$

which could also be derived from a random means model. In purely Bayesian inference, such as in [34], the prior grand mean η and prior precision λ are assumed known *a priori*.

Instead, an empirical Bayes approach would let the parameters η and λ be derived directly from the data. Therefore, this approach will outperform the classical sample mean because it relies on a richer model and includes the sample mean as the special case $\lambda = 0$.[3] The inadmissibility result found by Stein can be explained in a Bayesian framework by the fact that the sample mean corresponds to a diffuse prior, which is improper since it does not integrate to one. In that case, the Bayes rule need not be admissible.[4]

As discussed in Section II, optimal portfolio choice should be based on the predictive density function of the vector of future rates of return $\underline{r}$. With the informative prior (13), this predictive density function $p(\underline{r} \mid \underline{y},\Sigma,\lambda)$, conditional on Σ and λ, is multivariate normal, with mean

$$(14) \qquad E[\underline{r}] = (1 - w)\underline{Y} + w\underline{1}Y_0 ,$$

where $\quad w \equiv \dfrac{\lambda}{T + \lambda} \qquad Y_0 \equiv \underline{x}'\underline{Y} \equiv \dfrac{\underline{1}'\Sigma^{-1}}{\underline{1}'\Sigma^{-1}\underline{1}}\underline{Y},$

and covariance matrix

$$(15) \qquad V[\underline{r}] = \Sigma\left(1 + \frac{1}{T + \lambda}\right) + \frac{\lambda}{T(T + 1 + \lambda)} \frac{\underline{1}\underline{1}'}{\underline{1}'\Sigma^{-1}\underline{1}} .$$

The derivation is detailed in the Appendix. It is interesting to note that, after integration of η, the grand mean Y_0 happens to be the average return for the minimum variance portfolio.[5]

Zellner and Chetty [33] and Brown [11] studied the diffuse prior case $\lambda = 0$, where the moments reduce to

$$(16) \qquad E[\underline{r} \mid \underline{y}] = \underline{Y} \qquad V[\underline{r} \mid \underline{y},\Sigma] = \Sigma\left(1 + \frac{1}{T}\right).$$

For very large values of T, the correction due to estimation risk disappears: the moments $E[\underline{r}]$ and $V[\underline{r}]$ tend to the usual values $\underline{Y}$ and Σ, used in the certainty equivalence approach. But the richness of the empirical Bayes approach is that λ

[3] Efron and Morris [14] and Morris [28] describe this rationale in further detail.

[4] See, for instance, [4].

[5] Although not directly derived from a portfolio optimization process, the weights $\underline{x}$ minimize the variance of the portfolio subject to the condition that they sum to one.

was estimated from the data directly. The probability density function $p(\lambda \mid \underline{\mu},\eta,\Sigma)$ is a gamma distribution with mean at $(N + 2)/d$, where d is defined as $(\underline{\mu} - \underline{1}\eta)' \Sigma^{-1}(\underline{\mu} - \underline{1}\eta)$, and is replaced by its sample estimate $(\underline{Y} - \underline{1}Y_0)' \Sigma^{-1}(\underline{Y} - \underline{1}Y_0)$. The shrinkage coefficient is then

$$\hat{w} = \frac{N + 2}{(N + 2) + \left(\underline{Y} - Y_0\underline{1}\right)' T\Sigma^{-1}\left(\underline{Y} - Y_0\underline{1}\right)}, \tag{17}$$

which is a special form of (12).

In practice, Σ is unknown, and could be replaced, as in Zellner and Chetty [33], by

$$\hat{\Sigma} = \frac{T - 1}{T - N - 2} S, \tag{18}$$

where S is the usual unbiased sample covariate matrix. Substitution into (15) yields $\hat{V}[\underline{r}]$.

V. Practical Applications

The first goal of this paper was to demonstrate the inadmissibility of portfolio selection procedures based on sample means. The practical interest of this general result will now be illustrated by specific examples of potential gains from using shrinkage estimators. The performance of various estimation procedures should be measured by the loss of utility due to estimation error, averaged over repeated samples. Since this risk function is seldom analytically tractable, the natural procedure is to resort to simulation analysis.

Table 1 illustrates typical stock return data. These are sample estimates from stock market returns for seven major countries, calculated over a 60-month period. It is apparent that standard deviations are not too different across assets, and that they are very large relative to sample means. This is precisely a situation in which Bayes-Stein estimation is likely to help.

The parameters $\underline{\mu}$ and Σ were chosen equal to the estimates reported in Table 1. T independent vectors of returns were generated[6] from this distribution. For each sampling, the following estimators were computed:

1) Certainty Equivalence — $\underline{Y}, S$
2) Bayes Diffuse Prior — $\underline{Y}, \hat{V}[\underline{r}, \lambda = 0]$
3) Minimum Variance ($w = 1$) — $\underline{1}\, Y_0, \hat{V}[\underline{r}, \lambda \rightarrow \infty]$
4) Bayes-Stein ($\hat{w} = w(\underline{y},T)$) — $(1 - \hat{w})\underline{Y} + \hat{w}\,\underline{1}\, Y_0, \hat{V}[\underline{r}, \hat{\lambda}(\underline{y})]$.

Brown [11] has found the second estimator to be generally superior to the first one. The third estimator, advocated by Jobson et al. [19], is an extreme case of shrinkage, and has no formal justification in this context because the system is forced to be stationary. This choice, however, yields a particularly simple portfolio selection rule: for all utility functions, the optimal weights will be those of the minimum variance portfolio.

[6] Returns were generated by the IMSL subroutine GGNSM. All computations were performed in double-precision FORTRAN.

TABLE 1

Distribution Parameters for the Simulation Analysis

	Mean	Covariance Matrix						
Canada	1.287	42.18						
France	1.096	20.18	70.89					
Germany	0.501	10.88	21.58	25.51				
Japan	1.524	5.30	15.41	9.60	22.33			
Switzerl.	0.763	12.32	23.24	22.63	10.32	30.01		
U.K.	1.854	23.84	23.80	13.22	10.46	16.36	42.23	
U.S.	0.620	17.41	12.62	4.70	1.00	7.20	9.90	16.42
World	0.916	12.22						

Efficient Set Statistics

$c \equiv \underline{1}'S^{-1}\underline{1} = 0.11838$

$b \equiv \underline{Y}'S^{-1}\underline{1} = 0.0953$, thus $Y_0 = b/c = 0.805$

$a \equiv \underline{Y}'S^{-1}\underline{Y} = 0.15849$

$d(Y_0) \equiv (\underline{Y} - \underline{1}Y_0)'S^{-1}(\underline{Y} - \underline{1}Y_0) = 0.08171$

Notes: Dollar returns in percent per month. Sample period is January 1977—December 1981 ($T = 60$).

In order to find optimal weights, we had to define the investor's utility function. The negative exponential utility function was chosen here because of the existence of a closed-form solution for the optimal portfolio.[7] But the results should be robust to the choice of the utility function. A quadratic utility function gave essentially the same results, which are not reported here.

For each drawing k, the optimal portfolio was computed for each possible estimator, leading to different values of the derived expected utility $F_i = F[\underline{\hat{q}}(\underline{\mu}_i(\underline{y})) \mid \underline{\mu}, \Sigma]$, for $i = 1$ to 4. Repeating the experiment $K = 1000$ independent times, the empirical risk function was defined as the average loss of expected utility

$$\text{Risk}_i = \frac{F_{\text{MAX}} - F\left(\underline{\hat{\mu}}_i(\underline{y})\right)}{|F_{\text{MAX}}|}, \text{ with } F\left(\underline{\hat{\mu}}_i(\underline{y})\right) = \left(\frac{1}{K}\right)\sum_{k=1}^{K} F\left(\underline{\hat{\mu}}_i(\underline{y}_k)\right).$$

Expected utility was also expressed in risk-free equivalent return, as in (7). Finally, to study the effect of sample size, the previous operations were repeated for various values of T ranging from 25 to 200.

Figure 2 depicts the empirical risk functions, also reported in Table 2. Several features are apparent. First, the Bayes-Stein estimator has *always lower risk* than both the certainty equivalence and the Bayes diffuse prior estimators. The improvement is noticeable and significant. In risk-free equivalent, the gain over the diffuse prior case ranges from 8 percent ($T = 25$) to 2 percent ($T = 50$) to 0.2 percent ($T = 200$) per annum. The reason behind the superiority of the Bayes-Stein estimator is that the shrinkage factor $\hat{w}$ is directly derived from the data. For small sample sizes, large values of $\hat{w}$ indicate that portfolio analysis should not rely too much on the observed dispersion in sample means, given the large coefficients of variation of stock returns. But, of course, as the sample size

[7] The absolute risk aversion was chosen to be 1/(52.2% p.a.), as in Brown [11]. In annual terms, this implies that a 1 percent increase in the variance (10 percent in standard deviation) must be accompanied by an increase of about 1 percent in expected return.

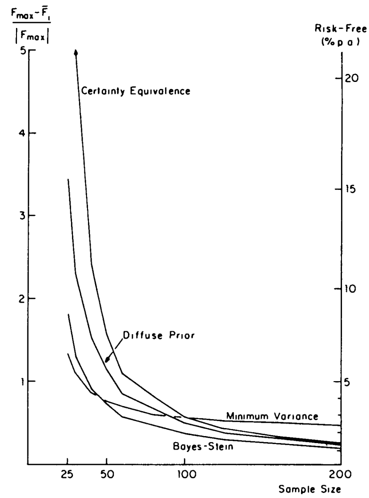

FIGURE 2

Empirical Risk Functions

TABLE 2

Empirical Risk Functions and Shrinkage Factors

Sample Size	Certainty Equivalence	Bayes Diffuse Prior	Minimum Variance	Bayes-Stein	Shrinkage Mean	Shrinkage Std. Dev.
25	1.5606	0.3452	0.1337	0.1815	0.5883	0.1564
30	0.4846	0.2294	0.1100	0.1301	0.5692	0.1464
40	0.2421	0 1526	0.0865	0.0916	0.5383	0.1427
50	0.1578	0.1137	0.0762	0.0722	0.5199	0.1386
60	0.1104	0.0856	0.0690	0.0576	0.5004	0.1365
80	0.0828	0.0683	0.0608	0.0473	0.4555	0.1352
100	0.0569	0.0493	0.0577	0.0375	0.4275	0.1221
125	0.0430	0.0385	0.0528	0.0308	0.3997	0.1171
160	0.0338	0.0310	0.0510	0.0256	0.3504	0.1073
200	0.0253	0.0236	0.0484	0.0205	0.3164	0.0906

Notes: Negative exponential utility function with absolute risk aversion of 1/(52.2% p.a.). Maximum expected utility given the true parameters is 0.99734.

increases, so does the estimated precision of sample means, and the shrinkage factor decreases, thus putting less weight on the informative prior relative to the data.

Next, the minimum-variance estimator performs well for small sample sizes, but is dominated for higher sample sizes. This is not astonishing: this strategy completely disregards any information contained in the sample means, which produces very good results for small samples, but is otherwise clearly inappropriate. For higher sample sizes, expected returns are more accurately estimated, and utility could be increased by taking into account the expected return of the portfolio. But this estimator would be particularly robust to nonstationarity. Finally, comparisons of the certainty equivalence and Bayes diffuse prior rules confirm the conclusions of Brown's study [11]: the Bayes diffuse prior uniformly dominates the classical rule.

Sections III and IV indicated that Bayes-Stein estimation should outperform the sample mean, whatever the true parameter values, and the simulation analysis indicated that the gains were substantial. Surely, these gains must be sensitive to the choice of the "true" parameter values, but it seems that these results provide conservative estimates of the gains from the Bayes-Stein estimator. Consider how different the means are in Table 1: expressed on a per annum basis, they vary from 6 percent to 22 percent. Insofar as this dispersion might be considered unrealistic, the simulation analysis will be biased *against* Bayes-Stein estimation. To take the case even further, the analysis was repeated with the highest mean changed from 22 percent to 44 percent per annum. The gains from shrinkage estimation were, on average, cut in half, but the Bayes-Stein estimator still uniformly dominated the sample mean.

VI. Conclusions

This paper studied the effect of estimation error on portfolio choice. Since parameter uncertainty implies a loss of investor utility, decision theory should be based on this loss viewed as a function of the estimator and of the true parameter values. A fundamental result of statistical theory is that the sample mean is an inadmissible estimator when the number of parameters is greater than two. (There exists another estimator that always yields lower expected loss in repeated samples.) This result stems from the summation of the effect of estimation error for each component into one single loss measure. Thus, the portfolio context is central to this result. The issue was also analyzed in an empirical Bayes framework, and this paper presented a simple Bayes-Stein estimator that should improve on the classical sample mean. Next, the extent of gains from Bayes-Stein estimation was illustrated by simulation analysis. The classical rule was always outperformed, and the gains were often substantial, in the range of a few percent per annum in risk-free equivalent return.

Numerous other applications of this technique are possible in finance. For instance, extensions to improved estimation of beta coefficients are straightforward. Also, Jorion [21] evaluated the out-of-sample performance of various estimators, based on actual stock return data, and found that shrinkage estimators significantly outperform the classical sample mean.

Appendix

Bayes-Stein Estimation

The problem is to find, as in Zellner and Chetty [33], the predictive distribution of future returns $\underline{r}$, conditional on the prior (13), the data $\underline{y} = (\underline{y}_1, \ldots, \underline{y}_t)$, on the covariance matrix Σ and on the scale factor λ

$$p(\underline{r} \mid \underline{y}, \Sigma, \lambda) = \int\int p(\underline{r}, \underline{\mu}, \eta \mid \underline{y}, \Sigma, \lambda) d\underline{\mu} d\eta \,. \tag{A.1}$$

When necessary, Σ and λ will be estimated from the conditional distribution. The joint density of $\underline{r}$, $\underline{\mu}$, and η is given by

$$p(\underline{r},\underline{\mu},\eta \mid \underline{y},\Sigma,\lambda) = p(\underline{r} \mid \underline{\mu},\eta,\Sigma,\lambda) \cdot p(\underline{\mu},\eta \mid \underline{y},\Sigma,\lambda)$$
$$\div\, p(\underline{r} \mid \underline{\mu},\Sigma) \cdot f(\underline{y} \mid \underline{\mu},\Sigma) p(\underline{\mu} \mid \eta,\lambda) p(\eta) .$$

With normality. the likelihood function of $\underline{y}_t$, given $\underline{\mu}$ and Σ, is

(A.2) $$f\left(\underline{y}_t \mid \underline{\mu},\Sigma\right) \div \exp\left[\left(-\frac{1}{2}\right)\left(\underline{y}_t - \underline{\mu}\right)'\Sigma^{-1}\left(\underline{y}_t - \underline{\mu}\right)\right] ,$$

and the density function of $\underline{\mu}$, given η and λ, is given by the informative prior

(A.3) $$p(\underline{\mu} \mid \eta,\lambda,\Sigma) \div \exp\left[\left(-\frac{1}{2}\right)(\underline{\mu} - \eta\underline{1})'\lambda\Sigma^{-1}(\underline{\mu} - \eta\underline{1})\right] .$$

Here, the λ parameter is a measure of the tightness of the prior; for λ tending to zero, the prior tends to a diffuse prior. The parameter η represents the unknown grand mean, which is given a diffuse prior. Instead of an informative prior on a model with constant means $\underline{\mu}$, (A.3) could also represent a model where the means $\underline{\mu}$ vary randomly around a common grand mean.

The predictive density function can then be written as

$$p(\underline{r},\underline{\mu},\eta \mid \underline{y},\Sigma,\lambda) \div \exp\left[\left(-\frac{1}{2}\right)G(\underline{r},\underline{\mu},\eta,\underline{y})\right], \text{ with}$$

$$G = (\underline{r} - \underline{\mu})'\Sigma^{-1}(\underline{r} - \underline{\mu}) + \sum_{t=1}^{T}\left(\underline{y}_t - \underline{\mu}\right)'\Sigma^{-1}\left(\underline{y}_t - \underline{\mu}\right)$$
$$+ (\underline{\mu} - \eta\underline{1})'\lambda\Sigma^{-1}(\underline{\mu} - \eta\underline{1}) .$$

After integration over η and $\underline{\mu}$, the predictive density can be shown to be normal with mean vector and covariance matrix as follows

(A.4) $$E[\underline{r}] = (1 - w)\underline{Y} + w\underline{1}Y_0,$$

w	= shrinkage factor	$w = \lambda/(T + \lambda)$,
$\underline{Y}$	= vector of sample means:	$\underline{Y} = (1/T)\sum_{t=1}^{T} \underline{y}_t$,
Y_0	= grand mean	$Y_0 = \underline{x}'\underline{Y}$,
$\underline{x}'$	= weights of min. var. portfolio:	$\underline{x}' = \underline{1}'\Sigma^{-1}/(\underline{1}'\Sigma^{-1}\underline{1})$.

(A.5) $$V[\underline{r}] = \Sigma\left(1 + \frac{1}{T + \lambda}\right) + \frac{\lambda}{T(T + 1 + \lambda)}\frac{\underline{1}\underline{1}'}{\underline{1}'\Sigma^{-1}\underline{1}} .$$

This covariance matrix has the following interpretation. The first term Σ represents the variation of $\underline{y}_t$ around the mean $\underline{\mu}$. The second term $\Sigma/(T + \lambda)$ is due to the uncertainty in the measure of the sample average $\underline{Y}$, whereas the third term corresponds to uncertainty in the common factor.

For T large, w tends to zero, $\underline{r}$ tends to $\underline{Y}$, and V tends to Σ. There is no estimation risk, and the sample means are accurate estimates of the expected returns. Similarly, for λ very small, w tends to zero, $E[\underline{r}]$ to $\underline{Y}$, and V to $\Sigma(1 + (1/T))$. Bayes-Stein estimation is useless here, and estimation risk, due to imprecise knowledge of expected returns, is impounded on V the usual way.

In contrast, for very large values of λ, w tends to one, $E[\underline{r}]$ to $\underline{1}\, Y_0$, and V to $\Sigma + (1/T)\, \underline{1}\, \underline{1}'(1/\underline{1}'\, \Sigma^{-1}\, \underline{1})$. Estimation of the means can be based only on the common weighted average Y_0, and the matrix added to Σ reflects uncertainty in this common mean.

References

[1] Barry, C. B. "Portfolio Analysis under Uncertain Means, Variances and Covariances." *Journal of Finance,* 29 (May 1974), 515-522.

[2] Bawa, V. S.; S. J. Brown; and R. W. Klein. *Estimation Risk and Optimal Portfolio Choice.* Amsterdam: North Holland (1979).

[3] Berger, J. "Minimax Estimation of a Multivariate Normal Mean under Polynomial Loss." *Journal of Multivariate Analysis,* 8 (June 1978), 173-180.

[4] ———. *Statistical Decision Theory.* New York: Springer-Verlag (1980).

[5] ———. "Selecting a Minimax Estimator of a Multivariate Normal Mean." *The Annals of Statistics,* 10 (March 1982), 81-92.

[6] Berger, J., and M. E. Bock. "Eliminating Singularities of Stein Type Estimators of Location Vectors." *Journal of the Royal Statistical Society,* 38 (1976), 166-170.

[7] Blume, M. "Betas and Their Regression Tendencies." *Journal of Finance,* 30 (June 1975), 785-795.

[8] Brown, L. D. "On the Admissibility of Invariant Estimators of One or More Location Parameters." *Annals of Mathematical Statistics,* 37 (Aug. 1966), 1087-1136.

[9] ———. "Estimation with Incompletely Specified Loss Functions." *Journal of the American Statistical Association,* 70 (June 1975), 417-427.

[10] ———. "A Heuristic Method for Determining Admissibility of Estimators—with Applications." *Annals of Statistics,* 7 (Sept. 1979), 960-994.

[11] Brown, S. J. "Optimal Portfolio Choice under Uncertainty: A Bayesian Approach." Ph.D. Diss., Univ. of Chicago (1976).

[12] ———. "The Effect of Estimation Risk on Capital Market Equilibrium." *Journal of Financial and Quantitative Analysis,* 14 (June 1979), 215-220.

[13] Dickenson, J. P. "The Reliability of Estimation Procedures in Portfolio Analysis." *Journal of Financial and Quantitative Analysis,* 9 (Sept. 1979), 447-462.

[14] Efron, B., and C. Morris. "Stein's Estimation Rule and its Competitors—An Empirical Bayes Approach." *Journal of the American Statistical Association,* 68 (March 1973), 117-130.

[15] ———. "Data Analysis Using Stein's Estimator and its Generalizations." *Journal of the American Statistical Association,* 70 (June 1975), 311-319.

[16] ———. "Families of Minimax Estimators of the Mean of a Multivariate Normal Distribution." *The Annals of Statistics,* 4 (Jan. 1976), 11-21.

[17] Frankfurter, G. M.; H. E. Phillips; and J. P. Seagle. "Portfolio Selection: the Effects of Uncertain Means, Variances and Covariances." *Journal of Financial and Quantitative Analysis,* 6 (Sept. 1971), 1251-1262.

[18] James, W., and C. Stein. "Estimation with Quadratic Loss." *Proceedings of the 4th Berkeley Symposium on Probability and Statistics 1.* Berkeley: Univ. of Calif. Press (1961), 361-379.

[19] Jobson, J. D.; B. Korkie; and V. Ratti. "Improved Estimation for Markowitz Portfolios using James-Stein Type Estimators." *Proceedings of the American Statistical Association,* Business and Economics Statistics Section, 41 (1979), 279-284.

[20] Jobson, J. D., and B. Korkie. "Estimation for Markowitz Efficient Portfolios." *Journal of the American Statistical Association,* 75 (Sept. 1980), 544-554.

[21] Jorion, P. "International Portfolio Diversification with Estimation Risk." *Journal of Business,* 58 (July 1985), 259-278.

[22] Klein, R. W., and V. S. Bawa. "The Effect of Estimation Risk on Optimal Portfolio Choice." *Journal of Financial Economics,* 3 (June 1976), 215-231.

[23] Lindley, D. V. "Discussion on Professor Stein's Paper." *Journal of Royal Statistical Society,* 24 (1962), 285-287.

[24] Lindley, D. V., and A. F. M. Smith. "Bayes Estimates for the Linear Model." *Journal of the Royal Statistical Society,* 34 (1972), 1-41.

[25] Markowitz, H. M. *Portfolio Selection: Efficient Diversification of Investments.* New York: Wiley and Sons (1959).

[26] Merton, R. C. "An Analytic Derivation of the Efficient Portfolio Frontier." *Journal of Financial and Quantitative Analysis,* 7 (Sept. 1972), 1851-1872.

[27] ________________. "On Estimating the Expected Return on the Market." *Journal of Financial Economics,* 8 (Dec. 1980), 323-361.

[28] Morris, C. N. "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association,* 78 (March 1983), 47-55.

[29] Stein, C. "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution." *Proceedings of the 3rd Berkeley Symposium on Probability and Statistics 1.* Berkeley: Univ. of Calif. Press (1955), 197-206.

[30] ________________. "Confidence Sets for the Mean of a Multivariate Normal Distribution." *Journal of the Royal Statistical Society,* 24 (1962), 265-296.

[31] Vasicek, O. "A Note on Using Cross-Sectional Information on Bayesian Estimation of Security Betas." *Journal of Finance,* 28 (Dec. 1973), 1233-1239.

[32] Zellner, A. *An Introduction to Bayesian Inference in Econometrics.* New York: Wiley and Sons (1971).

[33] Zellner, A., and V. K. Chetty. "Prediction and Decision Problems in Regression Models from the Bayesian Point of View." *Journal of the American Statistical Association,* 60 (June 1965), 608-615.

[34] Zellner, A., and W. Vandaele. "Bayes-Stein Estimators for *k*-Means Regression and Simultaneous Equations Models." In *Studies in Bayesian Econometrics and Statistics,* S. Fienberg and A. Zellner, eds. Amsterdam: North Holland (1974).