Journal of Financial Economics 3 (1976) 215-231. © North-Holland Publishing Company

THE EFFECT OF ESTIMATION RISK ON OPTIMAL PORTFOLIO CHOICE

Roger W. KLEIN and Vijay S. BAWA*

Bell Laboratories, Holmdel, N.J. 07733, U.S.A.

Received July 1975, revised version received December 1975

This paper determines the effect of estimation risk on optimal portfolio choice under uncertainty. In most realistic problems, the parameters of return distributions are unknown and are estimated using available economic data. Traditional analysis neglects estimation risk by treating the estimated parameters as if they were the true parameters to determine the optimal choice under uncertainty. We show that for normally distributed returns and 'non-informative' or 'invariant' priors, the admissible set of portfolios taking the estimation uncertainty into account is identical to that given by traditional analysis. However, as a result of estimation risk, the optimal portfolio choice differs from that obtained by traditional analysis. For other plausible priors, the admissible set, and consequently the optimal choice, is shown to differ from that in traditional analysis.

1. Introduction

The problem of portfolio choice under uncertainty has been traditionally viewed as a choice among alternative known probability distributions of returns. An individual chooses among them according to a consistent set of preferences. In practice, the distributions are assumed to belong to a certain family of distributions. Since the parameters characterizing each distribution are not ordinarily known, they are estimated using available economic data. The traditional approach is to use the distribution with the estimated parameters treated as if they were the true parameters to obtain an individual's optimal choice. By treating estimated parameters as true parameters, estimation risk is ignored. The purpose of this paper is to incorporate estimation risk directly into the decision process and determine its effect on optimal portfolio choice under uncertainty.

Under the Von Neumann-Morgenstern (1967) axioms, the optimal portfolio choice for an individual is a portfolio that maximizes the expected utility of returns, where the individual's utility function is determined uniquely, up to a

*We acknowledge helpful discussions with our colleague R.D. Willig. We thank members of the Holmdel Economic Theory Workshop at Bell Laboratories for stimulating discussions. We also thank the referee, Martin S. Geisel, and M.C. Jensen for helpful comments on an earlier draft.

positive linear transformation, by his preferences. The validity of the analysis is not affected by whether the probability distribution of portfolio returns is 'objective' or 'subjective', as long as it is completely specified. In the traditional analysis, one assumes that the estimated parameters of the distribution are the true parameters – thus completely specifying the distribution. In the Bayesian analysis, which explicitly considers estimation risk, we note that the Von Neumann–Morgenstern postulates allow us to solve the problem by using the predictive distribution of portfolio returns. We consider the case where the joint distribution of security returns is multivariate normal, determine the predictive distribution for several important prior distributions on the unknown parameters, and determine the effect of estimation risk on both the admissible set of portfolios and on the optimal portfolio choice.[1]

Section 2 outlines the traditional method of analysis by which the optimal portfolio choice is based on parameter (point) estimates, as well as the alternative Bayesian procedure that incorporates estimation risk directly into the decision process. The Bayesian method of analysis is not foreign to economists and has been applied to several non-portfolio choice type problems, see for example Zellner (1971), Zellner and Geisel (1968), Chow (1973), and Prescott (1972).

Section 3 examines optimal portfolio choice when there are an equal number of observations on each security's return and individuals have 'non-informative' or 'invariant' priors on the distribution's parameters. We show that the admissible set for all individuals is the same under the two procedures. However, more importantly, we show that the optimal choice for an individual, a point in this admissible set, will in general not be the same for the two procedures. As intuitively expected, for most risk-averse individuals, the optimal choice is likely to be a portfolio with lower expected portfolio return when the estimation risk is explicitly considered. An illustrative example is provided to demonstrate the extent to which decisions can differ under these two decision procedures.

Section 4 considers the portfolio choice problem for several general types of informative priors that one is likely to encounter. Two illustrative cases show that under the Bayesian procedure the admissible set can still be obtained by a very intuitive 'mean-variance' type selection rule. However, this rule differs (in a very intuitive way) from the mean-variance selection rule that one obtains under the traditional procedure. Our conclusions are stated in section 5.

2. The method of analysis

To explain the method of analyses, denote R_i as the (future) return to security i and X_i as the proportion of wealth invested in security i, $i = 1, \ldots, m$. Then, with $X' \equiv (X_1, \ldots, X_m)$ and $R' \equiv (R_1, \ldots, R_m)$, the return on portfolio X is

[1]The importance of estimation risk for portfolio choice is also discussed by Black and Treynor (1973).

$P_X = X'R$. With $U(\cdot)$ and $f(R \mid \theta)$ denoting the investor's utility function and the conditional distribution of R (conditioned on θ), respectively, the conditional expected utility of portfolio X is given by

$$\underset{R\mid\theta}{E}\ [U(X'R) \mid \theta] \equiv \int_R U(X'R) f(R \mid \theta)\,\mathrm{d}R. \tag{1}$$

In practical applications, θ is unknown, implying that $f(R \mid \theta)$ is not completely specified. In the traditional analysis one assumes that θ equals its estimate, $\hat{\theta}$. Then under this assumption, it is consistent with the Von Neumann–Morgenstern axioms for the investor's optimal portfolio to be given as the solution to the following:

$$\text{(I)} \qquad \underset{X}{\text{Max}}\ \underset{R\mid\theta}{E}\ [U(X'R) \mid \theta = \hat{\theta}] = \int_R U(X'R) f(R \mid \hat{\theta})\,\mathrm{d}R,$$

subject to

$$X \in C(X) = \{\Sigma X_i = 1,\ X_i \geqq 0\}.^2$$

Since the solution in (I) is based on $f(R \mid \hat{\theta})$, the distribution of R conditioned on $\theta = \hat{\theta}$, we call it the Estimated Conditional Solution, ECS.

It is important to note that in conditioning on $\theta = \hat{\theta}$, the ECS ignores estimation risk. In contrast, in the Bayesian procedure let

$$g(R) \equiv \underset{\theta}{E}\,[f(R \mid \theta)] \tag{2}$$

denote the unconditional (predictive) distribution of R, where in (2) the expectation is taken over θ with respect to the posterior distribution of θ, $p(\theta)$. Then, $g(R)$ completely specifies the probability distribution of R and, as will become clear below, incorporates estimation risk. In accordance with Von Neumann–Morgenstern axioms, the investor's optimal portfolio is now given as the solution to the following:

$$\text{(II)} \qquad \underset{X}{\text{Max}}\ \underset{\theta}{E}\left\{\underset{R\mid\theta}{E}\ [U(X'R) \mid \theta]\right\} = \underset{X}{\text{Max}} \int_R U(X'R) g(R)\,\mathrm{d}R,$$

subject to

$$X \in C(X).$$

[2]We note that the analysis in this paper is not effected if the set of feasible portfolios $C(X)$ is modified to include additional constraints of the type generally considered in the portfolio theory literature [e.g., Sharpe (1970)].

Since the method in (II) is based on the unconditional (predictive) distribution of R, we will call it the Unconditional Solution, UCS.

In this paper, we assume that the security returns R have a multivariate normal probability distribution, i.e., $f(R \mid \theta)$ is multivariate normal with $\theta = (\mu, \Sigma)$, where $\mu' \equiv (\mu_1, \ldots, \mu_m)$ denotes the mean vector and Σ denotes the variance-covariance matrix. Then, portfolio return $P_X \equiv X'R$ is normally distributed with mean $\mu_X = X'\mu$ and variance $\sigma_X^2 = X'\Sigma X$. Under these assumptions and others provided below, the optimization problems (I) and (II) reduce to the following:

$$\text{(I}'\text{)} \qquad \underset{X}{\text{Max}} \int_z U(\hat{\mu}_X + \hat{\sigma}_X \cdot z)\phi(z)\,\mathrm{d}z,$$

subject to

$$X \in C(X),$$

and

$$\text{(II}'\text{)} \qquad \underset{X}{\text{Max}} \int_y U(u_X^* + \sigma_X^* \cdot y)\eta(y)\,\mathrm{d}y,$$

subject to

$$X \in C(X).$$

In (I′), $\hat{\mu}$ and $\hat{\Sigma}$ denote the estimates of μ and Σ, respectively, $\hat{\mu}_X = X'\hat{\mu}$, $\hat{\sigma}_X = (X'\hat{\Sigma}X)^{\frac{1}{2}}$, and z is a random variable with standard normal density function $\phi(\cdot)$. In (II′), μ_X^* and σ_X^* are parameters that depend on the portfolio allocation X, and y is a random variable with p.d.f. $\eta(\cdot)$. We show that $\eta(\cdot)$ is either a standard normal or a standard t-distribution depending upon the prior employed. In the following sections, we obtain the solution to (II′) for several type of priors and compare it to the solution to (I′) to determine the effect of estimation risk on optimal portfolio choice.

3. Optimal portfolio choice: 'Non-informative' or 'invariant' priors

3.1. Theoretical analysis

In this section we consider the case of 'non-informative' or 'invariant' priors and obtain the optimal portfolio choice. We recall that the joint probability distribution of security returns R is assumed to be multivariate normal with mean μ and variance–covariance matrix Σ. Let r_{it} denote the (observed) return on security i, $i = 1, 2, \ldots, m$, at time t, $t = 1, 2, \ldots, T$, and let $\boldsymbol{r}' \equiv (r_{11}, r_{12}, \ldots, r_{1T}; r_{21}, \ldots, r_{2T}; \ldots; r_{m1}, \ldots, r_{mT})$:$1 \times mT$. Then the data, $D \equiv \boldsymbol{r}'$, is assumed to consist of observations drawn from a multivariate normal distribution with mean vector $(\mu_1, \ldots, \mu_1; \mu_2, \ldots, \mu_2; \ldots; \mu_m, \ldots, \mu_m)': 1 \times mT$ and variance–covariance matrix $\Sigma \otimes I$ (where I is a $T \times T$ identity matrix). Also, recall that for portfolio X, the portfolio return $P_X = X'R$ is normally distributed with mean $\mu_X = X'\mu$ and variance $\sigma_X^2 = X'\Sigma X$. Finally, we note that with $\theta = (\mu, \Sigma)$ denoting the unknown parameters and with $\propto$ implying 'proportional to' the 'non-informative' or 'invariant' prior $p_0(\theta)$ is given as

$$p_0(\theta) \propto |\Sigma|^{-(m+1)/2}. \tag{3}$$

We now define the parameters $\hat{\mu}_X$ and $\hat{\sigma}_X$ used in Theorem 1. Letting $\hat{\mu}' \equiv (\hat{\mu}_1, \ldots, \hat{\mu}_m)$, $\hat{\mu}_i \equiv \sum_{t=1}^{T} r_{it}/T$, $\hat{\mu}_X$ is given by

$$\hat{\mu}_X = X'\hat{\mu}. \tag{4}$$

Letting r be the $T \times m$ matrix of T observations on the m securities, with tth row given by $r_t' \equiv (r_{1t}, r_{2t}, \ldots, r_{mt})$, and $\mathbf{1}$ a $T \times 1$ column vector of ones, define

$$S \equiv (r - \mathbf{1}\hat{\mu}')'(r - \mathbf{1}\hat{\mu}'). \tag{5}$$

Then, $\hat{\sigma}_X^2$ is given as

$$\hat{\sigma}_X^2 \equiv X'SX/(T-m). \tag{6}$$

Before stating Theorem 1, it should be emphasized that $\hat{\mu}_X$ and $\hat{\sigma}_X^2$, which may be viewed as estimates of μ_X and σ_X^2, respectively, are defined prior to Theorem 1 only for convenience. The derivation of these estimates and their role in portfolio choice are determined by Theorem 1 as part of the decion problem.

Theorem 1. Under the assumption outlined above and for the non-informative prior given by (3), *the standardized unconditional (predictive) distribution* $\eta(\cdot)$ *used in* (II') *has a t-distribution with* $T-m$ *degrees of freedom. The parameters* μ_X^* *and* σ_X^* *are given by*

$$\mu_X^* = \hat{\mu}_X, \qquad \sigma_X^* = (1 + 1/T)^{\frac{1}{2}}\hat{\sigma}_X.$$

Theorem 1 is proved in the appendix.

Before using Theorem 1 to obtain the optimal portfolio choice, note that μ and Σ are independent in the prior and the prior for (μ, Σ) is 'non-informative' in that small changes in the data on security returns will change greatly the posterior p.d.f. for (μ, Σ).[3] In this sense, the data, as expressed through the likelihood function, plays the major role in determining the posterior distribution for these parameters.

Theorem 1 enables us to obtain the optimal portfolio choice under the UCS and compare it to the portfolio obtained using the ECS. Under the ECS (I′), the admissible set for all risk-averse individuals, as well as individuals with decreasing absolute risk-averse utility functions, is the well-known Markowitz–Tobin [Markowitz (1952, 1970), Tobin (1958)] mean-variance admissible boundary [see Bawa (1975) for details]. The admissible set for all individuals is the admissible set obtained by Bawa (1976). However, using our analysis, we see that with explicit consideration of estimation risk, the underlying distributions being compared are t-distributions with parameters μ_X^* and σ_X^*. Since $\mu_X^* = \hat{\mu}_X$ and $\sigma_X^* = (1+1/T)^{\frac{1}{2}}\hat{\sigma}_X$ and for t-distributions the admissible sets are still obtained with comparisons of μ_X^* and σ_X^*, it follows [see Bawa (1975, forthcoming) for the proof] that the admissible set for the same group of individuals is the same under the ECS and the UCS.

As stated above, consideration of estimation risk does not change the admissible set. However, more important, as a result of estimation risk, an individual's optimal choice (a member of the admissible set) is likely to differ from that obtained via the ECS. To show this result, we note that for all risk-averse individuals, the optimal portfolio choice lies on the Markowitz–Tobin mean–variance boundary. If we let $\sigma(\mu)$ denote this admissible boundary [see, for example, Markowitz (1970), Sharpe (1970) for a derivation of the boundary under several different sets of feasible sets $C(X)$], then it follows from Theorem 1 that the expected utility of return for the portfolio (on the admissible boundary) with expected return μ, denoted by $g(\mu)$, is given as

$$g(\mu) = \int_{-\infty}^{\infty} U[\mu + (1+1/T)^{\frac{1}{2}}\sigma(\mu)y]\eta_{T-m}(y)\,\mathrm{d}y, \tag{7}$$

where $\eta_{T-m}(\cdot)$ denotes the standard t-distribution with $(T-m)$ degrees of freedom. Thus, the optimal portfolio choice is obtained in two stages by (i) determining μ_* that maximizes $g(\mu)$ over feasible μ, and (ii) selecting the portfolio $X(\mu_*)$ corresponding to the optimal choice μ_*.

For increasing and concave utility functions ($U' > 0$, $U'' < 0$), μ_* is given by the First-Order Condition,

$$g'(\mu_*) \equiv \int U'[\mu_* + \gamma\sigma(\mu_*)y][1+\gamma\sigma'(\mu_*)y]\eta_{T-m}(y)\,\mathrm{d}y = 0, \tag{8}$$

[3]Formally, the marginal priors for μ and Σ are Jeffrey's invariant priors [Jeffrey (1961)]. The properties of these priors are summarized in Zellner (1971, pp. 41–53). Zellner (1971, pp. 42–44) and Box and Tiao (1973, pp. 25–41) explain the sense in which $p_0(\mu)\propto c$ is non-informative. Geisser (1965, pp. 602–603) provides an explanation of the sense in which $p_2(\Sigma)\propto|\Sigma|^{-(m+1)/2}$ is non-informative.

where $\gamma \equiv (1+1/T)^{\frac{1}{2}}$ and $\mu_* \equiv \mu_*(T)$ both depend on T. We note that the optimal value of μ under the ECS [solution to (I')] is given by $\bar{\mu} \equiv \mu_*(\infty)$. Since $\mu_*(T)$ depends on T through γ as well as the degrees of freedom $(T-m)$ of the t-distribution, for any finite T, $\mu_*(T) \neq \bar{\mu}$ for most increasing concave utility functions. Moreover, one might intuitively expect that as a result of increased (estimation) risk, $\mu_*(T)$ would be less than $\bar{\mu}$. Formally, letting

$$h(Y) \equiv U'(\mu+\gamma\sigma Y)[1+\gamma\sigma'(\mu)Y], \tag{9}$$

then the First-Order Condition (8) reduces to

$$E_Y h(Y) = 0.$$

It follows from Bawa (1975, theorem 13) that increasing the degrees of freedom $(T-m)$ of a t-distribution is equivalent to decreasing risk (for all increasing concave utility functions) as defined by Rothschild–Stiglitz (1971). Therefore, following the method of analysis in Rothschild–Stiglitz (1971, p. 67),

$$\mu_*(T) \lesseqgtr \bar{\mu}, \qquad \text{if} \quad h''(Y) \lesseqgtr 0. \tag{10}$$

Differentiating (9) and substituting $A \equiv -U''(w)/U'(w)$ and $R \equiv -wU''(w)/U'(w)$, the absolute risk-aversion and relative risk-aversion measures respectively yields

$$h''(Y) = \gamma^2\sigma^2(\mu)\left[\left(1-\frac{\mu\sigma'(\mu)}{\sigma(\mu)}\right)(A^2-A')-\frac{\sigma'(\mu)}{\sigma(\mu)}[A(1-R)+R']\right]U'. \tag{11}$$

Summarizing the above results, from (10) and (11), the following theorem is immediately established:

Theorem 2. For all risk-averse individuals, $\mu_*(T) \leqq \bar{\mu}$, if

$$\left[1-\mu\frac{\sigma'(\mu)}{\sigma(\mu)}\right](A^2-A')-\frac{\sigma'(\mu)}{\sigma(\mu)}[A(1-R)+R'] \leqq 0.$$

Theorem 2 immediately implies that $\mu_*(T) \leqq \bar{\mu}$, if

(i) $\quad U'''(\cdot) \leqq 0,$

or

(ii′) $\quad A' < 0, \quad$ (decreasing absolute risk-aversion), (ii″) $R \leqq 1$, (relative risk-aversion does not exceed one),

(ii‴) $R' \geqq 0$, (relative risk-aversion is non-decreasing), and

(ii⁗) $\mu\sigma'(\mu)/\sigma(\mu) \geqq 1$, over feasible μ.

This last condition $[\mu\sigma'(\mu)/\sigma(\mu) \geqq 1]$ always holds when the available securities include a riskless asset. Thus, for many (and perhaps most) cases, it appears that $\mu_*(T) \leqq \bar{\mu}$.

3.2. An illustrative example

The optimal portfolio choice $X[\mu_*(T)]$ is the optimal portfolio allocation for the point $\mu_*(T)$ on the mean-variance admissible boundary. To illustrate the quantitative effect of estimation risk on optimal portfolio choice, i.e., on $X[\{\mu_*(T)\}]-X(\bar{\mu})$, we consider the case where $m = 2$ and the utility function is the quadratic

$$U(y) \equiv 20y - y^2.\text{[4]} \tag{12}$$

Then, using notation defined earlier, it can be shown easily from the first-order condition (9) that

$$X_1^*(T) \equiv X_1[\mu_*(T)]$$

$$= \frac{10(\hat{\mu}_1-\hat{\mu}_2)-\hat{\mu}_2(\hat{\mu}_1-\hat{\mu}_2)-(\hat{\sigma}_{21}-\hat{\sigma}_{22})\alpha_T}{[(\hat{\mu}_1-\hat{\mu}_2)^2+(\sigma_{11}+2\hat{\sigma}_{21}+\hat{\sigma}_{22})\alpha_T]},$$

$$X_2^*(T) = 1 - X_1^*(T). \tag{13}$$

In (13), $\hat{\mu}_i$ is the estimate of μ_i, $\hat{\sigma}_{ij}$ is the estimate of σ_{ij} (an element of the variance-covariance matrix) and α_T is given by

$$\alpha_T = \left(\frac{T+1}{T}\right)\left(\frac{T-m}{T-m-2}\right). \tag{14}$$

For $i = 1, 2$ let X_i^e denote $X_i^*\,(T = \infty)$; then X_i^e is the optimal allocation under the ECS [i.e., (13) with $\alpha_{T=\infty} = 1$]. The optimal allocation under the UCS is given by $X_i^* \equiv X_i^*(T)$ from (13) with α_T defined by (14). To compare X_i^* with

[4]We recognize the problems associated with quadratic utility functions. We employ the quadratic to provide a simple but revealing example of the effect of estimation risk on optimal portfolio choice.

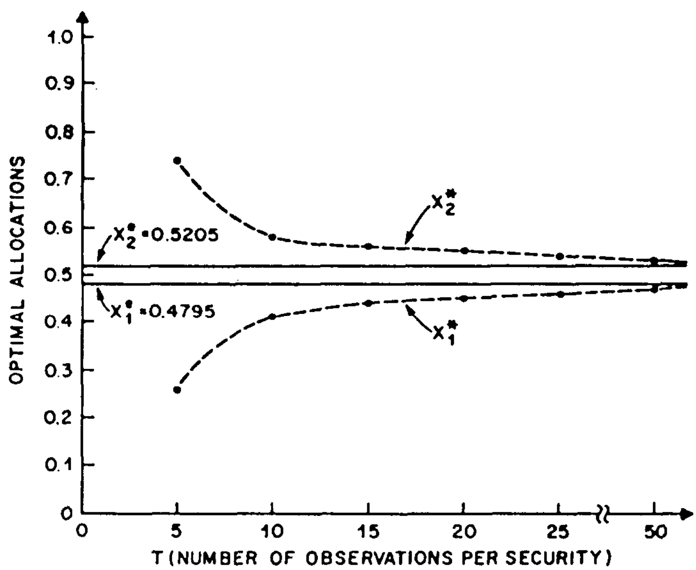
X_i^e, consider the special case

$$\hat{\mu}' = (12,\ 5), \qquad \hat{\Sigma} = \begin{pmatrix} 25 & 1 \\ 1 & 1 \end{pmatrix}, \tag{15}$$

for which

$$X_1^*(T) = 35/(49+24\alpha_T). \tag{16}$$

Note that the estimated standard deviation of security 1 is 5, which is not unrealistically high. We can now compare X_i^e with X_i^* when the estimates are based on $T = 5, 6, \ldots$, etc. observations. Accordingly, in fig. 1 we have plotted X_i^* and X_i^e as functions of T.



T	5	10	15	20	25	50	100	500
X_1^* (%)	25.85	41.57	44.16	45.25	45.85	46.95	47.46	47.71
$X_1^e - X_1^*$ (%)	22.10	6.38	3.79	2.70	2.10	1.00	.49	.24

THIS TWO SECURITY EXAMPLE ASSUMES A SPECIFIC QUADRATIC UTILITY FUNCTION AND A MULTIVARIATE NORMAL DISTRIBUTION OF SECURITY RETURNS. ESTIMATES OF THE UNKNOWN MEANS AND VARIANCE – COVARIANCE MATRIX ARE TAKEN AS GIVEN, WITH SECURITY 1 HAVING A HIGHER ESTIMATED MEAN AND VARIANCE THAN SECURITY 2. AS A FUNCTION OF THE SAMPLE SIZE (T), THIS FIGURE THEN SHOWS THE OPTIMAL ALLOCATIONS WHEN ESTIMATION RISK IS IGNORED (X_i^e, i = 1,2) AND WHEN IT IS TAKEN INTO ACCOUNT (X_i^*, i = 1,2).

Fig. 1

As shown in fig. 1, X_1^* is smaller than X_1^e and, as expected, monotonically approaches X_1^e as $T \to \infty$. Similarly, X_2^* is larger than X_2^e and monotonically approaches X_2^e as $T \to \infty$. Intuitively, fig. 1 shows that explicit consideration of estimation risk causes the investor to switch from the relatively 'riskier' security 1 to (the less risky) security 2. Fig. 1 and the accompanying table also show that the impact of estimation risk can be substantial. Estimation risk results in a 22 percent reduction in the investment in security 1 when the sample size $T = 5$, a 3 percent reduction when $T = 20$, and a 0.24 percent reduction even when $T = 500$. It is not uncommon to have 20 observations (or less) per security (5 years of quarterly data), so that estimation risk is likely to be an important consideration in practice. Moreover, in practice the number of securities is much larger than 2, implying that the degrees of freedom $(T-m)$ may be small. Consequently, the impact of estimation risk, via α_T in (14), is likely to be even more pronounced than shown here.

4. Optimal portfolio choice: Informative priors

In section 3 we showed that when the prior on the parameters (μ, Σ) is independent and non-informative, estimation risk does not effect the admissible set of portfolios, but does effect the optimal portfolio choice. In this section, we consider two illustrative cases of informative and dependent priors and show that estimation risk alters the admissible set of portfolios and hence again the optimal portfolio.

We first consider an extreme case in which the means of $(m-n)$ securities are known with certainty. The priors on the remaining n means are non-informative. Before proceeding, it should be noted that this case would be approximated if much more data were available for some securities than for others. Loosely speaking, therefore, this example can be interpreted as showing how the admissible set changes when much more information (prior or sample) is available for some securities than for others. For simplicity, in this example, we also assume that the variance-covariance matrix, Σ, is known to equal Σ^0.

Partition the vector of means μ and $V \equiv \Sigma^{-1}$ conformably as

$$\mu' \equiv [\mu'(1) \vdots \mu'(2)],\ \mu'(1){:}1 \times n,\ \mu'(2){:}1 \times m-n,$$

$$V \equiv \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}.$$

Given $\Sigma = \Sigma^0$ and $\mu(2) = \mu^0(2)$, the prior is

$$p_0[\mu(1) \mid \Sigma = \Sigma^0, \mu(2) = \mu^0(2)] \propto c, \quad \text{a constant.} \tag{17}$$

Under these assumptions, Theorem 3 obtains the standardized unconditional (predictive) distribution $\eta(y)$.

Theorem 3. Under the assumptions outlined in section 3 and for the prior $p_0(\cdot)$ given by (17), *the standardized unconditional (predictive) distribution $\eta(\cdot)$ used in* (II′) *is normal. Denoting $X(i)$, $i = 1, 2$, as a partition of X conformable with $\mu(i)$, the parameters μ_X^** and *σ_X^* are given as*

$$\mu_X^* \equiv X'(1)[\hat{\mu}(1) - V_{11}^{-1}V_{12}\{\mu^0(2) - \hat{\mu}(2)\}],$$

$$\sigma_X^* \equiv [X'(1)(V_{11}^{-1}/T)X(1) + X'\Sigma X]^{\frac{1}{2}}.$$

Theorem 3 is proved in the appendix.

Given the form of $\eta(\cdot)$ in Theorem 3, the admissible set can again be obtained by a 'mean-variance' type selection rule [see Bawa (1976, forthcoming) for details], where μ_X^* and σ_X^{*2} are the appropriate mean and variance measures, respectively. For example, for given μ_X^*, risk-averse individuals will minimize σ_X^*. However, since $\mu_X^* \neq X'(1)\hat{\mu}(1) + X'(2)\hat{\mu}(2) = X'\hat{\mu} = \hat{\mu}_X$ and $\sigma_X^* \neq \sigma_X$, decisons are no longer characterized by an estimated mean, $\hat{\mu}_X$, and the variance, σ_X^2, as they are in the ECS.

To intuitively explain Theorem 3, consider the bivariate case in which the portfolio consists of two securities, with returns r_{1t} and r_{2t}, $t = 1, \ldots, T$. Then, assuming the first security's mean, μ_1, is unknown while the second security's mean, μ_2, is known to equal μ_2^0 in the prior, it follows from Theorem 3 that

$$\mu_X^* = \hat{\mu}_X + X_1(\sigma_1/\sigma_2)\rho[\mu_2^0 - \hat{\mu}_2], \tag{18}$$

$$\sigma_X^* = [(1-\rho^2)\cdot X_1^2\cdot\sigma_1^2/T + \sigma_X^2]^{\frac{1}{2}}.$$

In (18), σ_1^2 and σ_2^2 are the variances of r_{1t} and r_{2t}, respectively, and ρ is the correlation coefficient between the two security returns. Turning first to μ_X^*, since the second mean, $\mu_2 = \mu_2^0$, is known, we are not directly interested in its estimate. However, if the data is misleading as to the second mean, and the means are correlated, then the data is misleading as to the first mean. For example, if $\mu_2^0 > \hat{\mu}_2$ (μ_2's estimate), then the data would understate the second mean (if one had employed the estimate rather than its known value, μ_2^0). If the returns are positively correlated ($\rho > 0$), then the data understates the first mean. Accordingly, the contribution of the first mean to expected portfolio return must be revised upward relative to its estimate (based solely on the data). The adjustment factor is given by

$$X_1(\sigma_1/\sigma_2)\rho(\mu_2^0 - \hat{\mu}_2). \tag{19}$$

Turning now to σ_X^{*2} in (18), note that it consists of two components, $(1-\rho^2)\cdot X_1^2\sigma_1^2/T$ and σ_X^2. The latter represents portfolio variance, while the former measures the impact of estimation risk. In a non-Bayesian analysis, σ_1^2/T would be the variance of μ_1's estimate, $\hat{\mu}_1$. Here, σ_1^2/T is the posterior variance of μ_1. It is essentially an additional risk element incurred as a result of not knowing μ_1.

If estimation risk (σ_1^2/T) is ignored, then a risk-averse individual would behave in closer accordance with Theorem 1. For a given appropriate mean measure, μ_X^*, portfolios would be chosen to minimize the variance, σ_X^2. If the first security has a low variance, this decision rule may result in a high proportion of the portfolio being invested in the first security, i.e., a high X_1. However, once estimation risk is taken into account, it is clear from (18) that less weight will be given to the first security. In this manner, the impact of estimation risk is reduced by reducing the proportion of the portfolio invested in that security whose mean is unknown. Intuitively, other things being equal, a risk-averse individual will want to invest more heavily in those securities about which he has the most knowledge.

To examine the implications of removing the independence assumption, assume that we are interested in comparing several mutual funds. Each consists of a set of securities; the funds differ in the proportions invested in the various securities. Assume that we only observe the overall return to the funds at time t, f_t^X, $t = 1, \ldots, T$, and that f_t^X is normally and independently distributed with mean μ_X and variance σ_X^2. Suppose that the mean, μ_X, is not independent of the variance, σ_X^2, in the prior. It may be, for example, that one initially believes that funds with a high mean have a high variance. In a simple formulation that captures this dependence, assume that

$$\mu_X = M(\sigma_X) + \varepsilon, \tag{20}$$

where $M(\cdot)$ is a known function (probably non-decreasing) and ε is a normally distributed error term with mean zero and known variance, σ_ε^2. Further, as stated above, for simplicity assume that σ_X^2 is known.

Employing these assumptions, the prior $p_0(\cdot)$ on the unknown mean μ_X is assumed to be normal with mean $M(\sigma_X)$ and variance σ_ε^2, i.e.,

$$p_0(\mu_X \mid \sigma_X) = N[M(\sigma_X), \sigma_\varepsilon^2]. \tag{21}$$

Then, the standardized predictive distribution $\eta(\cdot)$ is given by the following theorem.

Theorem 4. Under the assumption that portfolio returns at time t, f_t^X, $t = 1, \ldots, T$, are independently and normally distributed with mean μ_X and variance σ_X^2 and the prior $p_0(\cdot)$ given by (21), *the standardized unconditional* (*predictive*) *distribution used in* (II′) *is normal. The parameters* μ_X^* *and* σ_X^* *are given by*

$$\mu_X^* = [\hat{\mu}_X\sigma_\varepsilon^2 + M(\sigma_X)\sigma_X^2/T]/[\sigma_\varepsilon^2 + \sigma_X^2/T],$$

$$\sigma_X^* = \{[\sigma_\varepsilon^2\sigma_X^2/T]/[\sigma_\varepsilon^2 + \sigma_X^2/T] + \sigma_X^2\}^{\frac{1}{2}},$$

where

$$\hat{\mu}_X \equiv \sum_{t=1}^{T} f_t^X/T.$$

Theorem 4 is proved in the appendix.

From Theorem 4 it is readily apparent that the admissible set is again given by a 'mean-variance' selection rule [see Bawa (1976, forthcoming) for proof], where μ_X^* and σ_X^{*2} are the appropriate mean and variance measures, respectively. However, decisions cannot be characterized in terms of the fund's estimated mean, $\hat{\mu}_X$, and variance, σ_X^2, as they are in the ECS. For example, for risk-averse individuals [$U''(\cdot) < 0$], it is now possible that [assuming $M'(\sigma_X) > 0$] expected utility is an increasing function of σ_X. Consequently, for the same estimated mean, $\hat{\mu}_X$, a risk-averse individual, in choosing between two funds, might choose the fund with higher variance, σ_X^2. The problem here is that the relevant mean and variance measures are μ_X^* and σ_X^{*2}, not $\hat{\mu}_X$ and $\hat{\sigma}_X^2$. Since μ_X depends on σ_X in the prior, increases in σ_X may increase the relevant mean measure, μ_X^*, which in turn may increase expected utility. Conditions under which this holds depend on the specific form of the utility function.

This example also has implications for risk behavior. If individuals believe that μ_X and σ_X are positively related [$M'(\sigma_X) > 0$], which is not unreasonable, then 'observed' risk-seeking behavior under the ECS may be really risk-averse behavior in terms of the appropriate mean and variance measures when estimation risk is explicitly considered in the decision process.

5. Conclusions

In the theory of choice under uncertainty, optimal decisions are typically formulated in terms of certain parameters of underlying probability distributions. However, the parameters are seldom known in practical situations. It is therefore important to incorporate the estimation problem (parameter uncertainty) and the accompanying estimation risk into the general decision problem. The purpose of this paper has been to examine the effect of estimation risk on optimal portfolio choice.

We have shown for the optimal portfolio choice problem that for the normal distribution of returns and under 'non-informative' or 'invariant' priors, the admissible set is the same under the ECS and the UCS. However, as a result of estimation risk, the optimal member of this set under the UCS will differ from that under the ECS. For other types of priors that one is likely to encounter, under the UCS, the admissible set, and consequently optimal choice, was shown via illustrative examples to differ from that in the ECS. The method of analysis employed here could be extended to characterize capital market equilibrium when parameters are unknown. In this manner, the impact of estimation risk on capital market equilibrium can be taken into account. Such an analysis would integrate the Sharpe–Lintner–Mossin (1964, 1965, 1966) Capital Asset Pricing Model (that assumes that the parameters are known) with the econometric analysis [see, for example Jensen (1972a), Jensen (1972b) and references therein] done to estimate the unknown parameters. These issues will be examined in a separate paper.

Appendix: Proofs of theorems

Proof of Theorem 1. Expected utility (unconditional) is given by

$$\underset{R}{E}\,[U(X'R)] = \int_{X'R,\mu,\Sigma} U(X'R)h(X'R \mid \mu, \Sigma)p(\mu, \Sigma)d(\cdot),$$

where $h(\cdot)$ is the conditional p.d.f. of $X'R$. From Zellner (1971, pp. 233–236, 383–385), the predictive distribution is in the t-form, where with

$$\hat{\mu}' \equiv (\hat{\mu}_1, \ldots, \hat{\mu}_m), \qquad \hat{\mu}_i \equiv \sum_{t=1}^{T} r_{it}/T,$$

$$\mathbf{1}' \equiv (1, 1, \ldots, 1):1 \times m, \qquad S \equiv (r - \mathbf{1}\hat{\mu}')'(r - \mathbf{1}\hat{\mu}'),$$

$$\sigma_X^{*2} \equiv (1 + 1/T)X'SX/(T-m),$$

the variable $(X'R - X'\mu)/\sigma_X^*$ has a t-distribution with $T-m$ degrees of freedom. A detailed and alternative proof of this distributional result is available upon request from the authors. Letting $y \equiv (X'R - X'\hat{\mu})/\sigma_X^*$, the theorem immediately follows. E.O.P.

The following lemma will be useful in proving Theorems 3 and 4.

Lemma 1. Define

$$E[U(P_X) \mid \mu_X, \sigma_X] \equiv \int_{-\infty}^{\infty} U(\mu_X + \sigma_X \cdot z)\phi(z)\,\mathrm{d}z,$$

where z is a standard normal variable. Partition X and μ conformably as

$$X' \equiv [X'(1) \mid X'(2)], \qquad \mu' \equiv [\mu'(1) \mid \mu'(2)].$$

Assume that $\mu(1)$'s posterior p.d.f., given $\mu(2)$ and Σ, is multivariate normal with mean $\bar{\mu}(1)$ and variance–covariance matrix, Ω. Then, if Ω depends only on Σ,

$$E[U(P_X) \mid \mu(2), \Sigma] = \int_{-\infty}^{\infty} U[\mu_X^* + \sigma_X^* y^*]\phi(y^*)\, \mathrm{d}y^*,$$

where y^* is a standard normal variable, and

$$\mu_X^* \equiv X'(1)\bar{\mu}(1) + X'(2)\mu(2),$$

$$\sigma_X^* \equiv [X'(1)\Omega X'(1) + X'\Sigma X]^{\frac{1}{2}}.$$

Proof. Expected utility is given by

$$E[U(P_X) \mid \mu(2), \Sigma] = \int_{-\infty}^{\infty} U(\mu_X + \sigma_X \cdot z)\phi(z)p(\cdot)d(\cdot),$$

where $p(\cdot)$ is $\mu(1)$'s conditional (posterior) p.d.f. With H defined such that $H'H \equiv \Omega^{-1}$ and $Z^* \equiv H[\mu(1) - \bar{\mu}(1)]$, expected utility becomes

$$E[\cdot] = \int_{Z^*} \dots \int_z U[\mu_X^* + X'(1)H^{-1}Z^* + (X'\Sigma X)^{\frac{1}{2}} \cdot z]\phi(z)p_*(Z^*)d(\cdot),$$

where $p_*(\cdot)$ is the multivariate p.d.f. of the independent standard normal variables Z^*. Letting

$$y \equiv \mu_X^* + X'(1)H^{-1}Z^* + (X'\Sigma X)^{\frac{1}{2}} \cdot z,$$

$$E[\cdot] = \int_y U(y)f_y(y)\, \mathrm{d}y,$$

where y's p.d.f., $f_y(\cdot)$ is normal with mean μ_X^* and variance $\sigma_X^{*2} \equiv X'(1)\Omega X(1) + X'\Sigma X$. Therefore, with

$$y^* \equiv (y - \mu_X^*)/\sigma_X^*,$$

the lemma follows immediately. E.O.P.

Proof of Theorem 3. The likelihood function for $\mu(1)$ as a function of Σ^0, $\mu^0(2)$, and the data on security returns can be written as [see Zellner (1971, 380–383)]

$$L \propto \exp\left[-\tfrac{1}{2}\{[\mu'(1) - \hat{\mu}^{*\prime}(1)]T \cdot V_{11}[\mu(1) - \hat{\mu}^*(1)]\}\right],$$

where, with $\hat{\mu}(1)$ and $\hat{\mu}(2)$ denoting the appropriate vectors of sample means [estimates of $\mu(1)$ and $\mu(2)$, respectively],

$$\hat{\mu}^{*\prime}(1) \equiv \hat{\mu}'(1) - [\mu^0(2) - \hat{\mu}(2)]' V_{21} V_{11}^{-1}.$$

Therefore, the conditional posterior p.d.f. for $\mu(1)$ is multivariate normal with mean $\hat{\mu}^*(1)$ and variance-covariance matrix V_{11}^{-1}/T. Accordingly, the proof follows from Lemma 1. E.O.P.

Proof of Theorem 4. Given σ_X, expected utility is given by

$$E[U(P_X) \mid \sigma_X] = \int_{\mu_X} \int_z U(\mu_X + \sigma_X \cdot z)\phi(z) p(\mu_X \mid D)\, \mathrm{d}z \mathrm{d}\mu_X,$$

where z is standard normal and $p(\mu_X \mid D)$ is μ_X's posterior p.d.f. given data on fund returns over time, f_t^X, $t = 1, \ldots, T$. Since the prior for μ_X (given σ_X) is normal with variance σ_ε^2 and mean

$$M(\sigma_X),$$

the posterior p.d.f. of μ_X (given σ_X) will be normal [see Box–Tiao (1973, 15–18)], with mean

$$\mu_X^* \equiv \frac{\hat{\mu}_X \sigma_\varepsilon^2 + M(\sigma_X)\sigma_X^2/T}{\sigma_\varepsilon^2 + \sigma_X^2/T},$$

where

$$\hat{\mu}_X \equiv \sum_{t=1}^{T} f_t^X/T,$$

and variance

$$[\sigma_\varepsilon^2 \sigma_X^2/T]/[\sigma_\varepsilon^2 + \sigma_X^2/T].$$

The theorem now directly follows from Lemma 1.

References

Bawa, V.S., 1975, Optimal rules for ordering uncertain prospects, Journal of Financial Economics 2, 95–121.

Bawa, V.S., 1976, Admissible portfolios for all individuals, Journal of Finance, forthcoming.

Bawa, V.S., forthcoming, On optimal portfolio selection rules.

Black, F. and J.L. Treynor, 1973, How to use security analysis to improve portfolio selection, Journal of Business 46, 66–86.

Box, G.E.P. and G. Tiao, 1973, Bayesian inference in statistical analysis (Addison-Wesley, Reading, Mass).

Chow, G.C., 1973, Multiperiod predictions from stochastic difference equations by Bayesian methods, Econometrica 41, 109–118.

Geisser, S., 1975, A Bayes approach for comparing correlated estimates, Journal of the American Statistical Society 60, 602–607.

Jeffreys, H., 1961, Theory of probability, 3rd ed. (Oxford University Press, London).

Jensen, M.C., 1972a, Capital markets: Theory and evidence, The Bell Journal of Economics and Management Science 3, 357–398.

Jensen, M.C., 1972b, ed., Studies of the theory of capital markets (Praeger, New York).
Lintner, J., 1965, Security prices, risk and maximal gain from diversification, Journal of Finance 30, 587–615.
Markowitz, H., 1952, Portfolio selection, Journal of Finance 7, 77–91.
Markowitz, H., 1970, Portfolio selection: Efficient diversification of investments (Wiley, New York).
Mossin, J., 1966, Equilibrium in a capital market, Econometrica 34, 768–783.
Prescott, E.C., 1972, The multiperiod control problem under uncertainty, Econometrica 40, 1043–1057.
Rothschild, M. and J.E. Stiglitz, 1971, Increasing risk II: Its economic consequences, Journal of Economic Theory 3, 66–84.
Sharpe, W.E., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, Journal of Finance 29, 425–442.
Sharpe, W.F., 1970, Portfolio theory and capital markets (McGraw-Hill, New York).
Tobin, J., 1958, Liquidity preference as behavior towards risk, Review of Economic Studies 25, 65–86.
Von Neumann, J. and O. Morgenstein, 1967, Theory of games and economic behavior (Wiley, New York).
Zellner, A., 1971, An introduction to Bayesian inference in econometrics (Wiley, New York).
Zellner, A. and M. Geisel, 1968, Sensitivity of control to uncertainty and form of the criterion function, in: D.C. Watts, ed., The future of statistics (Academia Press, Oshkosh, Wisc.) 269–290.