A SIMULATION PROCEDURE FOR ESTIMATING BIAS IN WELL DIVERSIFIED PORTFOLIOS

by George M. Frankfurter
Syracuse University

and

Herbert E. Phillips
Temple University

ABSTRACT

The existence of a selection bias in applications of the portfolio selection models have previously been identified. The importance of this bias, in terms of the magnitude of its potential impact on portfolio selection, has never been demonstrated. Monte Carlo approaches are used in this paper in order to demonstrate that selection bias is more than a mere mathematical curiosity; the effects of this bias are very significant. Other insights are provided by the simulations. The point is made that these insights would be impossible to achieve without simulation.

I. BACKGROUND AND PURPOSE

In a related paper [4] the authors show analytically that, in models which follow the Markowitz tradition [7], the mere existence of random error in estimation is sufficient to introduce selection bias in applications. In application of the market model [9], for example, "those securities entering the first or high-beta portfolio would tend to have positive measurement errors in their β_j', and would induce positive bias in...the estimated portfolio risk coefficient" [1, p. 85].

In the companion paper [4] the authors establish, on purely theoretical grounds, the cause of selection bias, and the direction of its effects. These theoretical results were developed analytically and offered as explanation in part for a bias observed in some empirical work by Blum [2, pp. 7-8]. Blum studied the performance, over time of portfolios comprised of securities having similar estimated beta coefficients. For portfolios comprised of securities having the lowest beta estimates, a rise in the portfolio beta estimates one period later was invariably observed. Blum's empirical design, unfortunately, was neither rich nor flexible enough to provide any insight regarding the magnitude, consistency, or importance of the bias that he observed. A simulation approach would have had a clear advantage in this regard [5, p. 198].

In the present work, the authors re-examine various forms of bias which have previously been identified [1], [2]. By resort to Monte Carlo methods, a much richer framework for analysis is provided than any previously available [2], [4]. The purely theoretical [4], and the purely empirical [1], [2] results previously reported are varified here and placed in sharper focus. The efficacy of existing approaches for achieving efficient diversification is called into question.

II. PORTFOLIO SELECTION MODELS: AN IDEALIZED REPRESENTATION

Let n securities be considered for possible inclusion in an investment portfolio. A particular portfolio may be identified by a vector $\underline{x}$ whose i^{th} element ($i=1,2,\ldots,n$) represents the proportion of that portfolio which is invested in security i. Similarly, the expected returns are represented by means of a vector $\underline{\mu}$, and the variances and covariances by a variance-covariance matrix $\underline{\underline{\Sigma}}$.

Where rates of return are perfectly correlated:

$$\rho_{ij} = \sigma_{ij}(\sigma_i\sigma_j)^{-1}$$
$$= 1.00 \qquad \text{for all } ij$$

diversification cannot be achieved. But where rates of return are less than perfectly correlated:

$$\rho_{ij} < 1.00 \qquad \text{for some } ij$$

then, to the extent that such interrelationships can be known, this knowledge can be exploited to achieve diversification. Assuming that $\underline{\mu}$ and $\underline{\underline{\Sigma}}$ consist of known constants, the portfolio selection problem can be set up as follows

$$\text{Minimize} \qquad \phi = \underline{x}'\underline{\underline{\Sigma}}\,\underline{x} - \lambda\,\underline{\mu}'\underline{x}$$
$$\text{Subject to:} \quad \Sigma_{i=1}^{n} x_i = 1 \qquad (1)$$
$$x_i \geq 0 \quad i=1,2,\ldots,n$$

where λ is a constant, the coefficient of risk aversion, which measures the rate at which the investor is willing to exchange expectation for risk at the margin.

In an applications framework, the covariance approach (1) has been called laborious [7, p. 96-97]; questions have been raised [10] regarding the practicality of calling upon one's friendly analyist for, in effect, a nondiffuse prior [9] on each of $\{[3n + n^2]/2\}$ distinct inputs required for application of (1).[1] Markowitz suggested [7, p. 100], and Sharpe later popularized [10] a simplified model which requires fewer inputs than the covariance approach featured in (1). This simplification is made possible by assuming that the rates of return on "various securities are related only through common relationships with some basic underlying factor" [10, p. 281]. According to the market model formulation,

$$r_{it} = \alpha_i + \beta_i r_{mt} + \varepsilon_{mt} \quad (2)$$

where

r_{it} = the rate of return in period t on security i

r_{mt} = the rate of return in period t on a market portfolio

α_i = the rate of return on security i where $r_{mt} = 0$

β_i = the slope of a line showing the extent to which the rate of return on security i is affected by r_{mt}

ε_{it} = a normal random deviate, with zero mean and variance $Var(r_i) = Q_i$

The rate of return on a portfolio of risky assets can be viewed as the result of

1. a series of investments in n basic securities,

and

2. an investment in the index [10].

We exploit this interpretation and define

$$\beta_p = \Sigma_{i=1}^n x_i \beta_i \quad (3)$$

to be the weighted average response of r_{pt} to r_{mt}, where

r_{pt} = the rate of return in period t on portfolio p

β_p = the slope of a line showing the extent to which the rate of return on security i is affected by r_{mt}

The expected return and variance for any portfolio p can be written [10]:

$$E(r_{pt}) = \Sigma_{i=1}^n x_i \alpha_i + \beta_p E(r_{mt}) \quad (4)$$

$$Var(r_{pt}) = \Sigma_{i=1}^n x_i^2 \, Var(r_i) + \beta_p^2 \, Var(r_{mt}) \quad (5)$$

Thus, the objective function in (1) may be replaced by

$$\phi = \{\Sigma_{i=1}^n x_i^2 \, Var(r_i) + \beta_p^2 \, Var(r_{mt})\} - \lambda\{\Sigma_{i=1}^n x_i \alpha_i + \beta_p E(r_{mt})\} \quad (6)$$

An efficient algorithm for solving the market model formulation of the general portfolio selection problem is discussed in [10].

The market model provides more than a computational alternative to (1) [5]. According to an analogue [10, p. 282], portfolio risk can be dicotomized into two components:

1. a systematic component, $\beta_p^2 \, Var(r_{mt})$

and

2. a non-systematic component $\Sigma_{i=1}^n x_i^2 Var(r_i)$

According to theory [11], for properly diversified portfolios, all but the systematic component of risk will have been diversified away. In this paper we present evidence to the contrary, and in support of an entirely different proposition.

A further simplification of the portfolio selection problem is possible if the existence of a riskless asset is assumed [10, pp. 285-6].[2] Assuming that investors can borrow or lend at the same risk-free rate [10, p. 287], then, with just one exception, every point on a Markowitz efficient frontier [7] will be dominated by at least one point which lies on the locus of a straight line which is tangent to that frontier. The single exception, quite obviously, is provided by the point of tangency itself. Such a point identifies a unique collection of risky assets which we label S in honor of Sharpe [10].[3] The existence of a collection of risky assets which is efficient, unique, and invariant to investor utility, provides a simplification which we exploit in the Monte Carlo approach which follows.

III. MONTE CARLO APPROACH

Present approaches for selecting portfolios according to the mean-variance criteria do not account for the effect of error in estimation. In a related work it is argued that "the mere existence of an error component in the estimation functions for individual securities causes systematic bias in estimating characteristics of portfolios that are selected according to them" [4]. In this section a simulation model is described; the results of a preliminary application are described in the next section.

An analytic process is one which begins with data collection, and culminates when a decision is made. The object in this simulation is to replicate the analytic process for selecting portfolios which are efficient according to the mean-variance efficiency criteria [7]. The market model framework of Equation (2) is used [10]. In order to replicate the estimation functions for the market model parameters, α_i, β_i, and Q_i, we exploit the fact that the market model Equation (2) sets out a regression structure. Regression estimators have well known and tractable distributions [6, pp. 16-21]:

$$\tilde{\beta}_i \sim N\{\beta_i, Var(\tilde{\beta}_i)\} \quad (7)$$

$$\tilde{Q}_i \sim (Q_i/n)\chi_{n-2}^2 \quad (8)$$

$$\tilde{\alpha}_i = E(\tilde{r}_{it}) - \tilde{\beta}E(\tilde{r}_{mt}) \qquad (9)$$

when n = length of history, and the tilde distinguishes an estimator from a parameter. The expectation $E(\tilde{r}_{mt})$ and variance $Var(\tilde{r}_{mt})$ are treated as constants by the model, but the expectation

$$E(\tilde{r}_{it}) \sim N(\mu_i, Q_i/n) \qquad (10)$$

is a normal deviate. It follows from (9), and is otherwise well known [6, p. 21], that $\tilde{\alpha}$ and $\tilde{\beta}$ are jointly distributed, bivariate normal random variables. Monte Carlo methods are used to generate $\tilde{\alpha}$ and $\tilde{\beta}$ directly.

In order that the simulations be held within realistic bounds, the black box (i.e., simulation) parameters of (7), (8) and (10) are obtained by means of standard estimation procedures applied to actual data. A total of 108 observations on monthly rate of return were obtained for each of 760 securities, starting with February 1964. The black box parameters were estimated on the basis of the first 72 observations; the remaining 36 observations are held aside for the ex-post evaluations which may be undertaken at a later time. Values for the index were obtained by calculating rates of return on the geometric mean return for the 760 stocks.

In order that the number of simulation trials required to produce meaningful results be held to a manageable level, the existence of a risk-free asset is assumed. In each simulation trial, therefore, only one collection of risky assets, a proper subset of the universe of 760 securities, is efficient. This portfolio was identified by the symbol S in the previous section.[4]

The processes described by Equation (7), (8), (9) and (10) are easy to simulate by simple modification of any standardized normal random number generator. Programs of this sort are available at virtually any computer center. To generate a Chi-square random variable, it facilitates matters when a Chi-square program is available. Failing this, however, it should be noted that the sum of n standard normal deviates squared defines a Chi-square random variable.

Given a set of black box parameters, 100 repetitions for joint estimation of α_i and β_i were performed for each of the 760 securities in the universe, and for each security 100 independent repetitions of the process for generating Q_i were performed. The simulated random variables, $\tilde{\alpha}_i$, $\tilde{\beta}_i$, and $\tilde{Q}_i$, i=1,2,...,760, were set up, sequentially, in blocks. The blocks were stored as files on an auxiliary storage device. As a check on the system the expectations $E(\tilde{\alpha}_i)$, $E(\tilde{\beta}_i)$ and $E(\tilde{Q}_i)$ were calculated by averaging over the 100 simulated observations on each random variable in storage. The system showed excellent convergence. All the data inputs used in the simulation experiments that follow are on file, and are consistent with the data requirements that govern any regression structure [6].[5]

Black box parameters, which are consistent with the regression structure outlined by Equation (2), were obtained by standard statistical methods applied to 72 monthly observations on rate of return for each of 760 securities. A model was created to simulate this joint estimation process, using the black box parameters as inputs (i.e., to serve as expectations). With each application of this Monte Carlo process, a sample history consisting of six years of monthly observations on rate of return is, in effect, regenerated. The joint estimation process is run over and over again. The results of 100 replications of the estimation process are on file in the form of estimations $\tilde{\alpha}_i$, $\tilde{\beta}_i$, and $\tilde{Q}_i$ for each of 760 securities. The result of each such replication may serve as inputs for an application of the market model (2), just as a set of actual estimations might do. To limit the time and cost of this preliminary investigation, only 11 replications of the following experiment are performed; the results are presented below as preliminary results.

The effect of random estimation error on mean variance efficient portfolios is to be studied. The market model of Equation (2) is invoked, and the existence of a risk free rate is assumed. The risk free rate is set at $r_f = 0.0065$ on a monthly basis, which is equivalent to an annual rate of $r_f = 0.075$. For each of 11 simulated sample histories, the efficient frontier of portfolios consisting of risky assets is identified by solving

$$\begin{aligned} &\text{Minimize} && \phi = Var(r_{pt}) - \lambda E(r_{pt}) \\ &\text{Subject to:} && \Sigma_{i=1}^{n} x_i = 1 \qquad (7) \\ & && x_i \geq 0 \end{aligned}$$

A single point on each simulated efficient frontier is identified by finding the tangent of the line in E-V space [10] which passes through $r_f = .00625$. Note, this approach is independent of any assumptions about investor utility beyond that required for him to be risk averse. The results of these preliminary simulation results are presented in Table 1.

IV. EXPERIMENTAL RESULTS

In Table 1 various characteristics of efficient portfolios are described. A portfolio which is obtained on the basis of the black box parameters is said to be in "parameter space;" those which are obtained on the basis of simulated sample statistics are said to exist in "sample space." The results of 11 simulation trials are shown by rows 1 through 11 of the table; these portfolios were obtained in sample space. Expectations are formed by averaging over the 11 sample results for each characteristic. These expectations are displayed near the bottom of the table. The characteristics of the portfolios obtained on the basis of black box parameters are shown in the row labeled "parameter space."

In the companion paper [4] it was demonstrated analytically that mere random error in estimation would be sufficient to cause systematic bias. As a result of such bias, portfolios selected in accord with the mean-variance efficiency criteria would appear more desirable ex-ante than ex-post. Any difference (shown in the table) between the characteristics of efficient portfolios in sample space (rows 1 through 11 of the table) and that obtained in parameter space is the result of random sampling error, which is impossible to avoid in practice. The effects of such error, should be very closely approximated by these simulations.[6]

Referring to the table we see that for each efficient portfolio obtained in sample space, the expected rate of return shown in parameter space is overstated and the total risk is understated. On average, moreover, the degree of misstatement is in excess of 100 percent for both characteristics. Little wonder, therefore, that naive application of these approaches result, more than occasionally, in dissatisfied clients. The selection bias identified by [1] and explained by [4] is no mere mathematical curiousity; the effects of this bias are very significant. Further insight is provided by the simulation results shown in Table 1.

According to often stated gospel, where equilibrium is assumed, all but the systematic component of risk will have been diversified away by efficient portfolios. Assuming equilibrium conditions, of course, every efficient portfolio will consist of a proportionate share of the same collection of risky assets -- the market portfolio. Clearly, to assume general equilibrium is more restrictive than not to assume it. These results underscore the need for better justification than now exists for normative application of the Capital Asset Pricing logic [11].

Referring to that block of the table which is devoted to risk, we see that for a universe as large as 760 stocks efficient diversification will not, <u>in general</u>, result in the elimination of non-systematic risk. The non-systematic component of risk accounts for approximately 24 percent of total risk in this parameter space, and rises to nearly 40 percent, on average, in sample space. What explains this rise and what is its consequence?

From Equation (2) we see that the constants α_i and β_i are treated as independent constants by the market model. From Equation (9), however, it is clear that the regression estimates $\tilde{\alpha}_i$ and $\tilde{\beta}_i$ are not independent; they are jointly distributed and thus tied together by an error structure. The regression estimator $\tilde{Q}_i$ by contrast is independently distributed. Moreover, the process described by Equation (8) is fairly symmetric and peeked where there are 70 degrees of freedom or more, as in the present case. We see from the table that both systematic and non-systematic components of risk will be understated. Because of the joint relationship between $\tilde{\alpha}$ and $\tilde{\beta}$, however, and the model's thirst for high alpha and low beta, the systematic component of risk will be more seriously understated than the non-systematic component. Faced with a higher proportion of non-systematic risk in sample space than in parameter space, what should the program (7) do? The answer is obvious: the higher is the proportion of non-systematic risk to total risk, the greater is the potential advantage of Markowitz diversification [7]. The model (7) attempts to run down the non-systematic risk by bringing in more stocks. Notice, efficient portfolios are significantly larger in sample space than in parameter space. Moral: selection bias results not only in overstatement of expectation and in understatement of risk, but in superfluous diversification as well.

TABLE 1

Simulation No.	No. of Securities in Portfolio	Expected Return	Risk: Systematic	Risk: Non-System.	Risk: Total
1	45	4.41	1.46	1.55	3.01
2	46	3.89	1.59	1.14	2.73
3	51	4.01	1.66	1.44	3.10
4	51	3.61	1.77	1.25	3.02
5	57	3.94	2.56	1.38	3.94
6	38	4.17	1.96	1.27	3.23
7	49	3.98	2.51	1.18	3.69
8	51	3.88	1.54	1.25	2.79
9	48	3.98	1.91	1.24	3.15
10	43	3.82	2.31	1.06	3.37
11	45	4.45	2.57	1.38	3.95
Expectation	47.63	4.01	1.99	1.28	3.27
Parameter Space	39	1.72	5.26	1.66	6.92

FOOTNOTES

[1]The question that should be raised, of course, has to do with the practicality of calling upon security analysts to quantify their information, or to absorb quantitative information presented to them. The real importance of the "simplified" model which follows does not hinge on the fact that it requires only (3n + 2) distinct inputs [5].

[2]We wish to emphasize to the reader that the capital asset pricing model is <u>not</u> invoked in this paper.

[3]No general equilibrium conditions are implied. Assuming the existence of a riskless rate, the point S along any efficient frontier is invariant

to investor utility [10].

[4]Note that portfolio S is not a market portfolio in that, in the absence of any general equilibrium assumptions, it is in general not comprised of a weighted average of every stock in the universe.

[5]A minor point is worth noting. A data bank sufficient for 100 simulation trials was set up on auxiliary storage. The data need not have been created in this way, the system might have been programmed to generate the data as needed. This would not be cost effective, however. The cost of auxiliary storage is small and the cost of access is moderate. Random number generation, by contrast, is very expensive.

[6]Notice, moreover, that there is absolutely no way to approximate these effects without resort to simulation.

REFERENCES

1. Black, F., M.C. Jensen and M. Scholes, "The Capital Asset Pricing Model: Some Empirical Tests," in Studies in the Theory of Capital Markets, Michael Jensen (ed.) New York: Praeger (1972).
2. Blum, Marshall E., "On the Assessment of Risk," The Journal of Finance, XXVI (March 1971), 1-10.
3. Evans, John L., and Stephen H. Archer, "Diversification and the Reduction of Dispersion: An Empirical Analysis," The Journal of Finance, XXIII (December 1968) 761-768.
4. Frankfurter, George M., Herbert E. Phillips and John P. Seagle, "Bias in Estimating Portfolio Alpha and Beta Scores," The Review of Economics and Statistics, LVI (August 1974) 412-414.
5. Frankfurter, George M., Herbert E. Phillips and John P. Seagle, "Performance of the Sharpe Portfolio Selection Model: A Comparison," Journal of Financial and Quantitative Analysis, VI (June 1976) 191-204.
6. Johnston, J. Econometric Methods. New York: McGraw-Hill Book Company (1963).
7. Markowitz, Harry M., Portfolio Selection. New Haven: Yale University Press (1959).
8. Myers, Stewart C., "On the Use of β in Regulatory Proceedings: A Comment," The Bell Journal of Economics and Management Science, 3 (Autumn 1972) 622-627.
9. Raiffa, Howard and Robert Schlaifer, Applied Statistical Decision Theory. Boston: Harvard University Press (1961).
10. Sharpe, William F., "A Simplified Model for Portfolio Analysis," Management Science, 9 (Jan. 1963), 277-293.
11. Sharpe, William F., "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," Management Science, 3 (September 1964) 425-442.