Journal of Banking and Finance 15 (1991) 717–727. North-Holland

Bayesian and CAPM estimators of the means: Implications for portfolio selection

Philippe Jorion

Columbia University, New York, NY 10027, USA

Received August 1990, final version received December 1990

This paper compares active investment policies under three alternative models for estimating expected stock returns: the historical sample mean, a shrinkage or Bayesian estimator and a CAPM-based estimator. The out-of-sample performance of actively managed U.S. industry portfolios is analyzed for these three estimators over the period 1931 to 1987.

It is found that the classical method, based on historical means and covariances, leads to the worst forecasts and out-of-sample performance, and is generally outperformed by shrinkage estimators. An active portfolio with expected returns based on the CAPM produced the best results among all actively managed portfolios; this strategy, as we show, closely matches a simple buy-and-hold the market rule.

1. Introduction

The mean-variance framework developed by Markowitz (1959) for portfolio analysis is probably one of the most elegant and widely accepted concepts in financial theory. Unfortunately, its normative implications for portfolio selection are hindered by the difficulty of measuring the necessary inputs to the problem, namely expected returns, variances and covariances.

The effect of this uncertainty about underlying parameters, often called parameter uncertainty, has not always been appreciated in the literature, where unknown parameters are nearly always replaced by their sample values. An indication of this neglect is the continuous use of ex post mean-variance analysis to derive 'optimal' asset allocations in international finance.[1] The application of ex post mean-variance analysis is less frequent in domestic finance, where the CAPM, a model of market equilibrium, provides an alternative procedure for determining expected returns.

[1]This procedure was initiated by Grubel (1968), and used more recently by Adler and Dumas (1983). The latter rightly attract attention to problems with this approach.

0378-4266/91/$03.50

In fact, empirical studies based on simulation analyses[2] have cast some doubts on the reliability of classical estimation procedures. Most notably, Jobson and Korkie (1980) compare the true optimal weights to the distribution of sample values obtained in simulations, and conclude that classical estimation techniques perform very poorly. A common recurring problem is that small variations in sample means often lead to sizeable portfolio readjustments. This instability in optimal portfolio weights makes it difficult to recommend mean-variance analysis to portfolio managers.

It is not clear, however, that the sample means are the best estimators of future expected returns in the context of a portfolio. In fact, investors realize that sample values are affected by estimation error, and therefore may not be the best predictors of future parameters. Indeed investors may have prior information,[3] or choose a set of forecasts based on a model of market equilibrium, such as the CAPM.

This paper compares three classes of estimators: the classical sample mean, shrinkage estimators, and a CAPM-based estimator. Shrinkage estimators, also called Stein estimators, were introduced to portfolio analysis by Jobson and Korkie (1979), and have been studied by Jorion (1985), Dumas and Jacquillat (1990), and Grauer and Hakanson (1990). As Jorion (1986) shows, the multivariate nature of the portfolio optimization problem provides a sound statistical rationale for using shrinkage estimators in portfolio analysis.

Shrinkage estimators, however, are based on purely statistical arguments, and ignore risk/return tradeoffs that may be helpful in predicting stock returns. This is why this paper compares for the first time estimators based on the CAPM with other estimators.[4] Few studies seem to have emphasized the usefulness of the CAPM for providing unconditional forecasts of expected returns. Morgan (1978) compared predictors conditioned on a known market return, and found that the equally-weighted market outperforms the classical efficient portfolio outside the sample period. His results are similar to ours, but we focus on the unconditional situation where the future market return is not known, a much more difficult problem, but more realistic.

The objective of this study is to analyze the forecasting accuracy of various estimation methods based on actual data. We evaluate the ex post perfor-

[2]Frankfurter, Phillips and Seagle (1971) first compared portfolios selected according to mean-variance criteria to truly efficient portfolios. Jobson and Korkie (1980) studied the distribution of the weights of the portfolio that maximizes the ex ante Sharpe ratio.

[3]Zellner and Chetty (1965) first considered the problem of choosing optimal weights with uncertain parameters in a Bayesian framework. Other contributions have been made by Barry (1974), Brown (1979), Klein and Bawa (1976), among others.

[4]Whether the same risk/return tradeoff can be applied to international stock return data is an open question. One would have to assume away exchange rate risk and barriers to international investments.

mance of actual investment strategies, where optimal weights are constructed at each point in time on the basis of past information. This approach is practically oriented, and has the advantage of allowing for possible non-stationarity in the parameters, ruled out in simulation analyses. The disadvantage of this approach is that very long series are needed to obtain statistically significant results. This is why this paper uses the longest time period covered by the CRSP data base: industry indices compiled over the period 1926–1987. Whether historical means provide good out-of-sample forecasts for portfolio analysis is an important question, that can only be addressed using empirical analysis.

The paper is organized as follows. Section 2 reviews the definition and rationale of the estimators used in this study, while section 3 describes the data. In section 4 we evaluate the ability to forecast returns for each asset individually; however, a more relevant criterion may be the predictive performance in the context of a portfolio. The comparison methodology and the empirical results are presented in section 5. Finally, the last section summarizes the main points and compares the results with the existing literature.

2. Estimators for expected returns

We focus mainly on the estimation of expected returns, which, as indicated by Merton (1980), are much more difficult to estimate than variances or covariances, and critically influence the efficient frontier. With stationary independent normal returns, the classical – maximum likelihood – estimator for the expected rate of return, given the set of past observations y, is the sample mean:

$$E_{\mathrm{ML}}[r|y]=Y, \tag{1}$$

where r is the vector of future asset returns, and Y the vector of past sample means.

The classical approach relies exclusively on information contained in one series for each individual forecast. This may be inefficient, however, because pooling information contained in other series may result in better estimates of aggregate expected returns: assets with very high past returns, relative to the performance of the group, for instance, are likely to contain more positive estimation errors than others and, as a result, may not perform so well in the future. Given that the portfolio selection process aims at achieving the best portfolio allocation, and not necessarily at accurately forecasting individual asset returns, estimation risk can be statistically lowered by 'shrinking' sample averages toward a common value.[5] One such estimator, called a Bayes–Stein estimator, is constructed by Jorion (1986) as

$$E_{\mathrm{BS}}[r|y]=(1-w)Y+wY_0 1, \tag{2a}$$

$$Y_0=(1'S^{-1}Y)/(1'S^{-1}1), \tag{2b}$$

$$w=\lambda/(\lambda+T), \tag{2c}$$

$$\lambda=(N+2)(T-1)/[(Y-Y_0 1)'S^{-1}(Y-Y_0 1)(T-N-2)], \tag{2d}$$

where S is the usual sample covariance matrix, 1 is a vector of ones, N is the number of assets and T is the sample size. In this formulation, each sample mean Y is shrunk toward a common value, the mean of the minimum-variance portfolio Y_0. Other forms of shrinkage are possible, for instance shrinking toward the CAPM expected returns defined below, but the above specification will be the only one examined here. Because the shrinkage factor w is directly computed from the data – and not pre-specified – this class of estimators is also called 'empirical Bayes'. Although theoretically biased, this estimator has been shown to dominate the usual sample mean because of the multivariate nature of estimation risk.

For small sample sizes, usually below 60, the Bayes–Stein estimator was dominated in the simulation analysis by an extreme version, proposed by Jobson and Korkie (1979), who claim that there is no reliable evidence that common stock expected returns are different for small sample sizes. If all expected returns are set equal to the same value, the efficient frontier reduces to the minimum-variance portfolio, and

$$E_{\mathrm{MV}}[r|y]=Y_0 1. \tag{3}$$

Finally, the last estimator considered here is based on the Sharpe–Lintner CAPM, an economic model of market equilibrium: this estimator restricts expected returns to be solely a function of systematic risk. If we redefine all previous rates of returns as excess rates of return, the CAPM estimator can be written as

$$E_{\mathrm{CAPM}}[r|y]=\beta E[r_{\mathrm{M}}], \tag{4}$$

where β is the vector of coefficients from regressions of asset returns on the market, and $E[r_{\mathrm{M}}]$ the expected excess rate of return on the market. With the

[5]These estimators have been originally developed by James and Stein (1961). Zellner and Vandaele (1974) analyzed Bayesian methods of producing so-called Bayes–Stein estimators.

same assumptions as before, the beta coefficients can be estimated from their historical values. In addition, the expected return on the market can be forecast using the techniques developed by Merton (1980). The simplest model is based on a constant market risk premium, which because of risk aversion must be positive:

$$E_{\mathrm{CAPM}}[r|y]=\beta \max(0, Y_{\mathrm{M}}), \tag{5}$$

where Y_{M} is the past average excess return on the market. Results will be reported with the market proxied by the equally-weighted and value-weighted CRSP stock index.

3. Data description

The basic data consist of monthly returns on stocks listed on the NYSE during the period January 1926 to December 1987. All returns are measured in excess of the risk-free rate, using one-month Treasury Bill rates from the CRSP bond file. To consider a realistic portfolio allocation problem, the number of assets in the portfolio allocation problem was reduced to seven. Firms were first classified according to the first two digits of their SIC industry number; then the 65 industries were further aggregated into seven broad industry groups: energy, utilities, materials, consumer goods, capital equipment, trade and services, finance.[6] The classification is exhaustive, no firm is omitted. Each industry index was constructed as an equally-weighted average of all firms included in the industry without missing observations.

4. Forecasting asset returns

A first benchmark for the comparison of the four models is the *magnitude* of the forecast error. We used the following methodology: at each point in time, a forecast is generated on the basis of past information. A window corresponding to the last 60 months was selected, because this corresponds to a sample size commonly used in finance. Other windows were tried, with generally similar results. Thus the first forecast is created in December 1930, and compared to the subsequently realized value in January 1931. This yields a forecast error ε_t over the month of January. The procedure is repeated over the total perod 1931–1987, and the mean square forecast error $\mathrm{MSE}=(1/T)\sum_{t=1}^{T}\varepsilon_t^2$ is computed by averaging over the 684 'ex post' monthly observations.

Table 1 summarizes the forecasting performance of each model. The first impression is that the MSEs are not very different across models, with a

[6]Other grouping procedures, such as classifying on the basis of the first SIC number, do not alter the conclusions of our study.

Table 1

Forecasting performance of estimation methods: mean square forecast error. Out-of-sample period: January 1931–December 1987.[a]

	Mean-square forecast error				
				CAPM	
Asset	Mean	B–Stein	M-var	EW	VW
Energy	64.61	64.51	64.40	64.04*	64.10
Utilities	66.40	66.28	66.29	65.81*	65.89
Materials	47.74	47.54	47.39	47.21*	47.28
Consumer goods	73.17	73.11	73.03	72.54*	72.61
Capital equipment	64.44	64.35	64.29	63.91*	64.00
Services	72.09	71.80	71.58	71.23*	71.24
Finance	89.00	88.67	88.16*	88.28	88.11
Sum	477.44	476.28	475.14	473.02*	473.23
Of which, due to					
Bias	0.04	0.49	1.38	0.06	0.69
Variance	477.40	475.79	473.72	472.96	472.54

[a]Lowest MSE denoted by *. Forecasts based on past 60 months. Total number of observations is 684. With ε_t defined as the forecast error, the mean squared error $(1/T)\sum \varepsilon_t^2$ is decomposed into a bias component $\bar{\varepsilon}^2 = [(1/T)\sum \varepsilon_t]^2$ plus a variance component $(1/T)\sum (\varepsilon_t - \bar{\varepsilon})^2$.

difference of at most 2 percent. This simply reflects the fact that stock returns are very difficult to forecast, given that standard deviations are much larger than sample averages, typically by a factor of 10. Thus, when comparing MSEs, one cannot realistically expect differences of more than a few percent.

Looking at results across assets, table 1 shows that the classical estimator is always outperformed by other methods: sample means, which are constructed so as to minimize the *in-sample* variance, perform quite poorly outside the estimation period, and appear to be less robust than shrinkage estimators. Both the Bayes–Stein and the minimum-variance estimators improve on the sample mean, with the latter estimator giving the best results. At the bottom of the table, the MSE is decomposed into its bias and variance components. Relative to the classical mean, the shrinkage estimators are biased, but this systematic error is more than compensated by a low variance, so that overall, shrinkage estimators improve on the sample mean. Most interestingly, the CAPM based on the equally-weighted index (EW) performs rather well, with the lowest MSE for six out of seven assets. Whether better forecasts translate into better portfolio decisions is analyzed next.

5. Active investment strategies

The previous section provides some information on the forecasting ability of the four models under consideration; however, it may be argued that an investor is not necessarily interested in forecasting asset returns individually, but would rather prefer accurate estimates of optimal investment weights, or of the distribution parameters of optimal portfolio returns. This is why we also compare the estimators in the context of active investment management strategies. As before, portfolios were constructed each month on the basis of a moving window of 60 months.

With a riskless asset, the optimal portfolio maximizes the 'in-sample' Sharpe ratio, defined as the ratio of the portfolio expected excess return to its standard deviation, subject to the constraint that the proportions must sum to one. With short-sales allowed, the vector of optimal weights corresponding to estimator k is

$$q_k = S^{-1}E_k[r|y]/(1'S^{-1}E_k[r|y]). \tag{6}$$

With the observed large coefficients of variation for stock returns, using the same covariance matrix irrespective of the estimator for expected returns amounts to a second-order approximation, certainly negligible relative to the estimation error due to finite sample size.[7]

When the expected returns are derived from the CAPM, the optimal weights can be simplified to

$$q_{\mathrm{CAPM}} = (S^{-1}\beta)/(1'S^{-1}\beta), \tag{7}$$

and there is no need to forecast the expected return on the market, since it cancels out from the numerator and denominator. In addition, it should be noted that these weights should simplify to the market weights q_{m}, provided q_{m} is constant and no asset is omitted from the market, since in this situation β can be written as $(Sq_{\mathrm{m}})/\sigma_{\mathrm{m}}^2$. Here, given the exhaustive allocation of firms into industry groups, the only sources of differences between q_{m} and q_{CAPM} are rounding errors, or non-constant market weights.[8]

The procedure to compare active investment strategies is as follows. Each period, optimal weights are computed for each strategy; the ex post return for each portfolio is recorded the following month, at which time new optimal weights are generated from updated series. From these time-series of realized returns, we compute out-of-sample 'ex post' averages and variances, which can be compared to the average of their ex ante values.

[7]With monthly data, typical values of $\sigma[r]$ and $E[r]$ are 10% and 1%, respectively. Therefore, when computing variances, the term $E[r^2]$ will dominate $E[r]^2$ by a factor of 100 to 1. The choice of different estimators for $E[r]$ should therefore not affect very much the covariance matrix.

[8]For the equally-weighted CAPM, the market weights are nearly constant over time, except for the addition and deletion of firms, and the optimal weights should be very close to the market weights.

Table 2
Performance of alternative investment strategies. Out-of-sample period: January 1931–December 1987.[a]

Strategy	Mean		Standard deviation		Sharpe	
	Ex ante average	Ex post	Ex ante average	Ex post	Ex ante average	Ex post
Risk-free asset		0.287		0		0
Passive						
Equal weighted	1.024	1.095	7.897	7.788	0.130	0.141
Value weighted	0.619	0.671	5.757	5.703	0.108	0.118
Optimized portfolios						
Classical	3.446	0.940	10.737	17.933	0.321	0.052
Bayes–Stein	1.556	0.743	7.941	9.776	0.196	0.076
Minimum variance	0.693	0.891	5.418	7.160	0.128	0.124
CAPM (EW)	1.109	0.988	7.385	7.354	0.150	0.134
CAPM (VW)	0.789	0.723	6.606	7.417	0.119	0.097

[a]Ex post statistics are out-of-sample measures; ex ante statistics are averages of in-sample measures. Forecasts based on past 60 months. Number of ex post observations is 684. Sharpe measure defined as the ratio of mean return over standard deviation. All returns are excess returns in percent per month, except those for the risk-free rate.

Ex post performance can be measured either by total or systematic risk. As Jobson and Korkie (1981) show, the Sharpe statistic is well behaved in small samples, although not very powerful; on the other hand, the Treynor test statistic performs much worse in simulation analyses, even for large sample sizes. We will focus primarily on the Sharpe measure, first because of its acceptable small sample properties, and second because it is the most relevant measure here, since the optimal portfolio is constructed so as to maximize this statistic, ex ante.

For the period 1931–1987, table 2 reports in-sample and out-of-sample statistics for optimal portfolios based on the four previous estimation methods. These active strategies were compared to a passive rule of buy-and-hold the market, which is proxied by the CRSP equally-weighted and value-weighted indices.

From table 2 it appears that in-sample and out-of-sample means differ markedly, with ex post returns falling short of their expectations for optimized portfolios. In contrast, standard deviations seem easier to forecast. The classical portfolio, supposedly the best if estimation risk is ignored, turns out to be the *worst* out-of-sample, because it relies heavily on unstable past means.

On the other hand, shrinkage portfolios, based on more conservative estimators, lead to much better results. The Bayes–Stein strategy clearly improves on the classical approach, but not as much as the minimum-variance portfolio, which leads to excellent results, considering that its sole

Table 3

Tests of equal portfolio performance. Out-of-sample period: January 1931–December 1987.[a]

	Paired Z-tests		
	Against model		
Model	Classical	Bayes–Stein	Min-var
Bayes–Stein	−1.27		
Min-var	−1.56	−1.30	
CAPM (EW)	−1.67*	−1.42	−0.40
CAPM (VW)	−0.89	−0.51	1.30

[a]Significance at the 10% level denoted by *. Portfolio performance refers to investment strategies described in table 2. The null hypothesis is that of equal ex ante portfolio performance for two models. A negative entry indicates that the row model outperforms the column model. Asymptotic z-statistics based on the tests developed by Jobson and Korkie (1981).

objective is to minimize variances. These results are consistent with previous simulations that have shown the superiority of the minimum-variance portfolio over Bayes–Stein portfolios for small sample sizes. Note that this portfolio does achieve its objective of minimizing the total portfolio risk, since it displays the lowest 'ex post' risk among all actively managed portfolios. This again indicates that historical covariance matrices contain useful information which can be used to forecast future risk measures, but that historical averages are unreliable guides to the future.

Finally, the EW CAPM portfolio appears to be the most successful of the four active strategies, even though the simplest scheme was adopted for estimating betas; this success can be traced to reasonable and stable estimates of expected returns. In addition, the CAPM portfolios behave quite similarly to the market indices on which they are based. The good results for the minimum-variance portfolio can also be interpreted as a particular application of the CAPM, where all beta coefficients are set equal to one; the fact that the CAPM-based portfolio still beats the minimum-variance portfolio indicates that there is some stability in the beta estimates and that those are cross-sectionally related to expected returns.

To check whether these performance measures are statistically different, we tested the hypothesis that the Sharpe ratios were identical across strategies, using the methodology developed by Jobson and Korkie (1981). The statistics are presented in table 3. In spite of the reported low power of the tests, the table indicates that the classical portfolio is significantly outperformed by the EW CAPM strategy. It should be noted that in economic terms, the differences in the reported Sharpe ratios are sizeable. For instance, comparing the classical and EW CAPM active portfolios, the Sharpe values of 0.052 and 0.134 translate into annual excess returns of 4.8% and 12.5%, respectively, for a level of volatility set equal to that of the market. Thus, after controlling for risk, the EW CAPM portfolio outperforms the classical portfolio by 7.7% per annum, which is barely significant in statistical terms but economically very significant.

6. Conclusion

These results cast serious doubts on the applicability of classical ex post mean-variance analysis. They show that ex post performances of actively managed portfolios systematically fall short of expectations. Because of the difficulty of estimating expected returns, active optimized strategies can perform worse than simple buy-and-hold investments.

The results in this paper demonstrate the benefits from using shrinkage estimators, and are in broad agreement with previous research by Jorion (1985), using international stock data, and Dumas and Jacquillat (1990), using foreign exchange data. Different conclusions, however, were reached for an international stock return data set by Grauer and Hakansson (1990). They report a situation where shrinkage estimators are outperformed by classical rules. These contrasting results are probably due to differences in the data set and in the methodology, the most significant of which is the imposition of short-sales restrictions. Further research is necessary to ascertain the exact reasons for these differences. All studies based on actual data, however, are plagued by the low signal-to-noise ratio that is typical of stock returns; as a result, very long time-series are required to achieve statistically significant results.

In conclusion, among the active strategies analyzed here, the classical strategy led to the poorest results. The sample means seem to contain little useful information in the context of portfolio selection. In contrast, shrinkage estimators brought a noticeable improvement over optimized portfolios based on past means. The best results, however, were attained with expected returns based on the CAPM, which highlights the central message of the CAPM: hold the market.

References

Adler, M. and B. Dumas, 1983, International portfolio choice and corporation finance: A survey, Journal of Finance 38, 925–984.

Barry, C.B., 1974, Portfolio analysis under uncertain means, variances and covariances, Journal of Finance 29, 515–522.

Brown, S.J., 1979, The effect of estimation risk on capital market equilibrium, Journal of Financial and Quantitative Analysis 14, 215–220.

Dumas, B. and B. Jacquillat, 1990, Performance of currency portfolios chosen by a Bayesian technique: 1967–1985, Journal of Banking and Finance 14, 539–558.

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, 1251-1262.

Grauer, R. and N. Hakansson, 1990, Stein and CAPM estimators of the means in portfolio choice: A case of unsuccess, Manuscript.

Grubel, H.G., 1968, Internationally diversified portfolios: Welfare gains and capital flows, American Economic Review 58, 1299–1314.

James, W. and C. Stein, 1961, Estimation with quadratic loss, Proceedings of the 4th Berkeley Symposium on Probability and Statistics 1 (University of California Press, Berkeley) 361–379.

Jobson, J.D., B. Korkie and V. Ratti, 1979, Improved estimation for Markowitz portfolios using James–Stein type estimators, Proceedings of the American Statistical Association, Business and Economics Statistics Section, 279–284.

Jobson, J.D. and B. Korkie, 1980, Estimation for Markowitz efficient portfolios, Journal of American Statistical Association 75, 544–554.

Jobson, J.D. and B. Korkie, 1981, Performance hypothesis testing with the Sharpe and Treynor measures, Journal of Finance 36, 888–908.

Jorion, P., 1986, Bayes–Stein estimation for portfolio analysis, Journal of Financial and Quantitative Analysis 21, 279–292.

Jorion, P., 1985, International portfolio diversification with estimation risk, Journal of Business 58, 259–278.

Klein, R.W. and V.S. Bawa, 1976, The effect of estimation risk on optimal portfolio choice, Journal of Financial Economics 3, 215–231.

Markowitz, H.M., 1959, Portfolio selection: Efficient diversification of investments (Wiley, New York).

Merton, R.C., 1980, On estimating the expected return on the market, Journal of Financial Economics 8, 323–361.

Morgan, I.G., 1978, Market proxies and the conditional prediction of returns, Journal of Financial and Quantitative Analysis 13, 943–946.

Zellner, A. and V.K. Chetty, 1965, Prediction and decision problems in regression models from the Bayesian point of view, Journal of the American Statistical Association 60, 608–615.

Zellner, A. and W. Vandaele, 1974, Bayes–Stein estimators for k-means regression and simultaneous equations models, in: S. Feinberg and A. Zellner, eds., Studies in Bayesian econometrics and statistics (North-Holland, Amsterdam) 629–653.