International Portfolio Diversification with Estimation Risk

Author(s): Philippe Jorion

Source: *The Journal of Business*, Jul., 1985, Vol. 58, No. 3 (Jul., 1985), pp. 259-278

Published by: The University of Chicago Press

Stable URL: https://www.jstor.org/stable/2352997

Philippe Jorion
Columbia University

International Portfolio Diversification with Estimation Risk*

Previous studies of international portfolio diversification have relied on ex post mean-variance analysis without considering the problem of estimation risk. Typically, past averages are substituted for expected returns and no allowance is made for the uncertainty inherent in these parameter values. First I demonstrate the shortcomings of such an approach, and second I investigate another method of estimating expected returns initially proposed by Stein. By shrinking the sample averages toward a common mean, I find that the out-of-sample performance of the optimal portfolio is substantially increased. One of the implications of this method is that the classical conclusions vastly overestimate the possible gains in average returns; instead, benefits from diversification are more likely to accrue from a reduction in risk.

I. Introduction

International portfolio diversification has long been advocated as a way of enhancing average returns while reducing portfolio risk for the investor who considers diversifying into foreign securities. This proposition, however, relies on the assumption that the required inputs to the classical mean-variance analysis are known with certainty. Typically, expected returns, variances, and covariances are simply replaced by their ex post sample values and the optimal portfolio is derived without mentioning the uncertainty inherent in these parameter values. But the rational investor should take this uncertainty into account when forming expectations, and probably will consider estimators that are less subject to estimation error than the classical sample mean. This paper investigates alternative estimators of expected returns and their implications for the alleged gains from diversification.

An important observation is the crucial influence that errors in estimating expected returns have on portfolio analysis. A closer exami-

* This paper draws on part of my dissertation at the University of Chicago. I am grateful to Arnold Zellner, Michael Mussa, George Constantinides, and especially John Bilson, for useful discussions and comments. Numerous improvements were also suggested by the referee. This research received the generous financial support of the College Interuniversitaire d'Etudes Doctorales dans les Sciences du Management.

(*Journal of Business*, 1985, vol. 58, no. 3)

0021-9398/85/5803-0002$01.50

nation of optimal portfolios based on past sample mean returns as estimates of expected returns reveals a sharp deterioration of performance measures outside the sample period used to calculate means, and an instability of the optimum weights. Both problems can be traced to wide fluctuations in sample means. On the other hand, uncertainty in variances and covariances is not as critical because they are more precisely estimated.

Because the problem lies in the measure of expected returns, estimators other than the sample mean should be explored. One alternative is the class of shrinkage estimators proposed by Stein (1955). These have been introduced to portfolio analysis by Jobson, Korkie, and Ratti (1979),[1] and further developed by Jorion (1984). Basically, each asset mean should be "shrunk" toward a common value, say the world mean, for predicting expected returns. Jorion (1984) showed that shrinkage estimators appropriately deal with this parameter uncertainty issue in portfolio analysis: in theory, they uniformly reduce estimation risk relative to the usual sample mean. In this paper I argue that Stein estimators improve out-of-sample performance substantially and have important implications for portfolio selection.

Obviously the issue raised is very general and is not restricted to the analysis of international equity investments. I chose this context to illustrate the argument because of the proliferation of papers based on ex post mean-variance analysis in international finance. This phenomenon is caused partly by the difficulty of generalizing domestic asset pricing models to the international environment,[2] which has led empirical research to rely heavily on the risk-reward framework first proposed by Markowitz (1959).[3]

The paper is organized as follows. Section II highlights the major defects of the classical application of mean-variance analysis, while Section III suggests that most of the mentioned problems can be traced to the imprecision in the measure of expected returns. In Section IV, I present an alternative class of estimators, less subject to estimation error than the sample mean, and study their impact on the geometry of

1. Jobson et al. (1979) argue that the shrinkage should be extreme: all expected returns should be set equal to the grand mean, irrespective of the sample size. This conclusion is difficult to reconcile with the generally accepted trade-off between risk and expected return, unless all stocks fall within the same risk class.

2. International asset pricing models have been proposed by Solnik (1974*a*, 1974*b*), Grauer, Litzenberger, and Stehle (1976), Losq (1979), Fama and Farber (1979), Stultz (1981). A comprehensive survey by Adler and Dumas (1983) offers a clarifying analysis of the theoretical issues. But the few empirical tests performed to date have not been powerful enough to distinguish between alternative hypotheses. See for instance Solnik (1977) or Stehle (1977).

3. A short list of studies based on ex post mean-variance analysis includes Grubel (1968), Levy and Sarnat (1970, 1975, 1979*a*, 1979*b*), Levy (1981), Logue (1982), Solnik (1982), and most recently Adler and Dumas (1983).

the efficient set. These shrinkage estimators indicate that most of the gains in average returns claimed by studies of international portfolio diversification are illusory but that the risk reduction factor is highly significant. In Section V, I simulate an active investment strategy that confirms the previous proposition. The paper ends with a summary and conclusions.

II. Major Defects of the Classical Mean-Variance Analysis

Mean-variance analysis, as advocated by a number of authors, has serious shortcomings that are too often ignored. This section illustrates the major defects of the classical approach. For illustrative purposes, I tracked the optimal portfolios found by Grubel (1968) and Levy and Sarnat (1970); table 1 reports their composition, as well as in-sample and out-of-sample performances. By construction, these portfolios maximize the in-sample Sharpe ratio, subject to a no-short-selling constraint. The Sharpe index is a standard performance measure, defined as the ratio of the portfolio expected return in excess of the risk-free rate over the portfolio standard deviation.

Perhaps the most serious defect of the classical approach is the poor out-of-sample performance of the optimal portfolios. Performance measures always deteriorate substantially outside the sample period, and the supposedly optimal choice is sometimes dominated by a naive method. In table 1, for instance, the Grubel portfolio is outperformed by a simple equally weighted index.

Another problem is the instability of the optimal portfolio: the proportions allocated to each asset are extremely sensitive to variations in expected returns, and adding a few observations may change the portfolio distribution completely. Also, optimal portfolios are not necessarily well diversified.[4] Often a corner solution appears, where most of the investments are zero and large proportions are assigned to countries with relatively small capital markets and high average returns.

These problems are major barriers to the practical application of mean-variance analysis and warrant a closer examination of the nature of estimation risk.

III. The Problem of Estimating Expected Returns

One of the required inputs to mean-variance analysis is a set of expected rates of returns for the assets under study. Assuming stationary multivariate normal returns, expected values are traditionally proxied

4. In the Levy and Sarnat (1970) study, only seven countries out of the initial group of 28 enter the optimal portfolio. In general, when short sales are not allowed, most of the weights found by quadratic programming are exactly zero.

TABLE 1 **Comparison of Selected Portfolios**

A. Composition in Percentages

	Levy and Sarnat 1951–67 (28 Countries)	Grubel 59.01–66.12 (11 Countries)	Equally Weighted Index	World Index (December 1960)
Austria	7.0	N.A.	11.1	.1
Australia	.0	37.3	11.1	2.7
Canada	.0	15.9	11.1	2.5
France	.0	2.7	11.1	2.6
Italy	.0	1.5	11.1	2.6
Japan	16.7	7.0	11.1	3.4
Mexico	4.3	N.A.	.0	.0
New Zealand	6.3	N.A.	.0	.0
South Africa	12.5	15.4	11.1	.8
U.K.	.0	7.6	11.1	9.4
U.S.A.	41.0	12.5	11.1	65.7
Venezuela	11.7	N.A.	.0	.0

B. Annualized Performance Measures

	Mean	SD	Sharpe	Mean	SD	Sharpe	Mean	SD	Sharpe	Mean	SD	Sharpe
In sample	12.5	29.1	.323	8.8	22.1	.262	...	...	...	...	...	...
Out of sample:												
Jan. 68–Dec. 69	6.9	29.7	.132	12.9	40.8	.243	11.4	27.8	.302	4.7	41.8	.041
Jan. 70–Dec. 71	11.1	40.8	.199	3.0	49.0	.001	5.6	34.7	.075	6.7	48.5	.076
Jan. 72–Dec. 73	18.1	40.0	.376	14.2	40.5	.277	15.8	42.1	.304	4.0	50.4	.019
Jan. 74–Dec. 75	−.3	66.2	−.050	−1.7	94.1	−.050	−2.4	73.2	−.074	.8	68.6	−.032
Jan. 76–Dec. 77	7.9	33.2	.149	3.1	41.3	.002	2.1	36.6	−.024	7.8	34.3	.141
Jan. 68–Dec. 77	8.7	43.6	.132	6.3	56.4	.059	6.5	45.4	.077	4.8	49.2	.037

NOTE.—Returns expressed in percentage per year. Optimal portfolios chosen on the basis of a constant U.S. risk-free rate of 3% per year. Data source for out of sample performance measure is Capital International and excludes Mexico, New Zealand, and Venezuela.

by sample means.[5] The question is, How useful are these sample means for predicting future returns?

This issue is addressed using actual data from seven major equity markets.[6] Data sources and computations of national equity returns converted into dollars[7] are detailed in Appendix A. The period 1971–83 was the most recent for which these data were available.

First, I shall focus on summary statistics for the data, reported in the first column of table 2. The essential characteristic of these stock return series is a very small mean relative to the standard deviation. Even after averaging over 144 data points, I am unable to reject the hypothesis that the expected return is equal to zero at the usual 5% confidence level: in the case of the United States, for instance, the relevant t-statistic is $0.62/(4.47/\sqrt{144}) = 1.66$, which is still low. Of course, this does not imply that the true population values are zero, but wide fluctuations in sample means will occur, and it should even be possible to measure negative values of sample averages for some sampling period.[8]

Next, I adopt the following framework for studying the predictive performance of average returns: for each observation t, the past sample mean is defined as the moving average of the n observations in the "window" $t - n$ to $t - 1$. Define r_{jt} as the rate of return for country j at time t and $\bar{r}_{jt-1}$ the average over the n months, ending at $t - 1$.

In this case, the shorter the window n, the longer the out-of-sample period available with which to analyze the behavior of the estimate. With 13 years of data, a 12-month window, for example, leaves 12 years for the analysis of forecasting performance. Results will be reported for 12-month and 60-month windows; fortunately, the main conclusions are not sensitive to the choice of the window.

The analysis of forecasting efficiency of past averages is based on Fama's (1976) methodology: the realized return for country j at time t is regressed against the past average

$$r_{jt} = \beta_j \bar{r}_{jt-1} + \epsilon_{jt}. \tag{1}$$

5. Note that the sample mean is an efficient estimator only if both the expected value and the variance are constant through time.

6. The fact that I limit my analysis to seven major markets permits a reduced computational burden while still covering more than 89% of the world index market value as of December 1976.

7. I adopt the viewpoint of an investor whose consumption basket is denominated in dollars. Hence prices should be measured in dollars. This domestic currency approximation is probably closest to reality for U.S. investors.

8. Merton (1980) also observed that estimating expected returns from time series of realized stock returns data is very difficult. He suggested using prior information based on the capital asset pricing model and decomposed the expected return on the market into two components: a risk-free rate and an excess return, maybe time varying. With a suitable international asset pricing model, this approach could be extended to national equity returns.

TABLE 2 **Forecasting Efficiency of Past Averages**
(Standard Deviations in Parentheses)

	Equity Returns (Mean)	OLS Regression $r_{jt} = \beta_j \bar{r}_{jt-1} + e_{jt}$: b	$t(b = 1)$	$R^2 = 1 - \frac{SSE}{SSY}$	D-W
	Twelve-Month Averages: Jan 72–Dec 83 (144 Observations)				
Canada	.81 (6.23)	.261 (.247)	−2.99**	.0077	2.10
France	.65 (7.58)	.114 (.256)	−3.46**	.0014	1.75
Germany	.85 (5.45)	.213 (.293)	−2.69**	.0037	2.01
Japan	1.33 (5.62)	.571** (.180)	−2.38**	.0658	1.83
Switzerland	.75 (5.74)	.194 (.287)	−2.81**	.0032	1.89
U.K.	.66 (8.24)	.099 (.267)	−3.37**	.0010	1.74
U.S.A.	.62 (4.47)	.300 (.247)	−2.83**	.0102	2.01
	Five-Year Averages: Jan 76–Dec 83 (96 Observations)				
Canada	1.13 (6.66)	.616 (.715)	−.54	.0078	2.05
France	.58 (7.55)	−.685 (.934)	−1.80	.0056	1.88
Germany	.71 (5.07)	.355 (.578)	−1.12	.0040	2.19
Japan	1.36 (5.06)	.771 (.438)	−.52	.0317	1.93
Switzerland	.73 (4.95)	.270 (.571)	−1.28	.0023	1.86
U.K.	1.19 (6.55)	.429 (.533)	−1.07	.0068	1.80
U.S.A.	.98 (4.04)	1.277 (.603)	.46	.0451	2.03

NOTE.—Dollar returns in percentage per month. Standard errors in parentheses. Standard deviations below mean returns refer to the row series, not to the standard error of the estimated means. The t-statistic tests the hypothesis that the slope coefficient is unity.
** Significant at 1% level.

If the average is an unbiased forecast, the coefficient β should be unity.[9] Using ordinary least squares in this framework amounts to minimizing the sum of squared forecast errors.

9. The model was chosen without a constant. The series of equity returns are so "noisy," in terms of signal extraction, that autocorrelation can hardly be detected. Actually, one might consider a model with constant expected returns. In this case, the intercept would reflect this constant term, and the coefficient of the past average would be approximately zero. If the intercept is eliminated, however, I hope that the *level* of returns—rather than variations—will be explained by past averages. In addition, with slowly varying expected returns, the moving average should capture some of the nonstationarity in the series.

Regression results are presented in the rest of table 2. In addition to the poor predictive power, revealed by low values of R^2, note that, with the possible exception of Japan, all slope coefficients are close to zero. They are also significantly far from unity: from a Bayesian perspective, the probability that the parameters are equal to one is always less than 4%. This suggests that 1-year historical averages are poor if not useless predictors of future returns in univariate models. The 5-year averages fare somewhat better, but the lower t-values are mainly due to a smaller number of observations for the out-of-sample test period, which entails larger standard errors.

Alternatively, these results can be explained by the problem of errors in the variables. Assume for the moment that the expected return μ and the variance σ^2 are truly constant for country j. Then the moving average will measure μ with error, the variance of the error being (σ^2/n), where n is the length of the moving window: $\bar{r}_{t-1} = \mu + \eta_t$. Even if the true relationship is $r_t = \beta\,\mu + \epsilon_t$, with $\beta = 1$, the regression coefficient b will be inconsistent:

$$\text{plim } b = \frac{\beta}{\left(1 + \frac{\sigma^2}{\mu^2}\frac{1}{n}\right)}.$$

The bias for stock return data is far from trivial. For instance, with $n = 12$ and typical values of the coefficient of variation $\sigma/\mu = 10$ and 5, plim $b = .11$ and .32, respectively. With a longer period for the moving average $n = 60$, these values increase to .38 and .71, still far from unity. As a matter of fact, six out of seven estimated betas increase when n goes from 12 to 60.[10]

Whatever the interpretation, these results clearly emphasize the importance of estimation risk and the imprecision in measured expected returns.

IV. Estimator Choice and the Efficient Set

Having recognized that the distribution parameters are unknown, one can explore other estimators than the sample mean, which is, after all, just one functional form among many. To forecast the return of each country, one need not be restricted to past information on the country only: benefits might accrue from pooling the data from all countries. For example, a better predictor could be obtained by "shrinking" the country sample mean toward a common value, which is less likely to be affected by extreme observations than the country mean. This in a

10. The U.S. coefficient is greater than one, but this need not invalidate the argument: the expression for the bias is only valid asymptotically.

nutshell is the basic idea behind Stein's "shrinkage" estimators.[11] In this section I discuss Stein estimators and their implications for the efficient set.

Estimation risk can be studied in the Bayesian framework proposed by Zellner and Chetty (1965), Klein and Bawa (1976), and Brown (1979). The approach can be summarized as follows. Assume the sample means have been transformed to be independent, normally distributed with equal variance:[12] $\bar{\mathbf{r}} \simeq N(\boldsymbol{\mu}, \sigma^2 I)$. Then in the strict Bayesian approach the data are combined with, say, a normal prior on expected returns, $\boldsymbol{\mu} \simeq N(r_0\mathbf{1}, \tau^2 I)$, so that the posterior mean becomes

$$\mu_j|\bar{\mathbf{r}}, \quad \text{prior} = \frac{\sigma^2}{\tau^2 + \sigma^2} r_0 + \frac{\tau^2}{\tau^2 + \sigma^2} \bar{r}_j. \tag{2}$$

One of the arguments often raised against this approach is that the choice of the prior parameters r_0 and τ^2 seems arbitrary. On the other hand, restricting ourselves to an "objective," or diffuse, prior is equivalent to setting τ^2 to a very large value, in which case the posterior mean equals the usual sample mean; unfortunately, in that situation, the corrections for estimation risk are not significant for most practical purposes.[13]

Instead, Stein estimators only specify a *class* of priors, as before, but the prior parameters r_0 and τ^2 are estimated directly from the data. Such an approach, also called "empirical Bayes,"[14] will outperform the classical sample mean because it relies on a richer model and includes the sample mean as a special case. Thus the general form for Stein estimators is

$$\bar{r}_j(\hat{w}) = \hat{w}\hat{r}_0 + (1 - \hat{w})\bar{r}_j, \tag{3}$$

where $\hat{r}_0$ and $\hat{w}$ are estimated from the data. Here the country average is "shrunk" toward a grand mean $\hat{r}_0$. At this point many specifications are possible, but the emphasis is on the consideration of a wider range of models rather than on the particular model selected.

11. These estimators, developed by Stein (1955) and James and Stein (1961), have been hailed by Lindley (1962) as "the most important statistical idea of the decade." Stein bolstered his argument with Bayesian considerations, developed more fully by Zellner and Vandaele (1974). Efron and Morris (1973, 1975) present applications of Stein's estimators.

12. If the covariance matrix Σ is known, the means should be transformed to $P\bar{\mathbf{r}}$, where P is defined from $T\Sigma^{-1} = P'P$. It can be verified that since the covariance matrix of means is Σ/T, where T is the number of observations, the covariance matrix of the transformed means is the identity matrix and the prior parameter σ^2 is unity.

13. With a diffuse prior, the mean of the predictive density function is the usual sample average; the covariance matrix is multiplied by a constant, which is typically small: for a sample size of 100, and 10 assets, the correction factor is 1.14. In addition, with a riskless asset, Brown (1979) has shown that the optimal portfolio *weights* do not differ from the classical ones.

14. The empirical Bayes approach is clearly exposed in Morris (1983).

Generally, shrinkage estimators will have lower estimation risk than the sample mean where estimation risk is measured as the average loss of investor's utility in repeated samples. As Jorion (1984) shows, this result is a direct application of Stein's work to the portfolio context: the efficiency gain derives from the simultaneous estimation of many parameters. I shall illustrate this by an example. Assume I want to assess the effect of measurement errors in the grand mean. If my aim were to forecast expected returns individually, then obviously the choice of the grand mean would be essential for the precision of the forecast. However, what matters here is the influence of the grand mean on the optimum *weights*, and this influence is likely to be much less important than in the previous case, because the relative ranking of expected returns is unaffected by changes in the grand mean. In fact, for the negative exponential utility function, it can be shown that the choice of the grand mean is even completely irrelevant for optimal portfolio choice.[15]

Therefore, the fundamental element in Stein estimators is the shrinkage factor, which is derived from the chosen specification for the prior. For instance, Jorion (1984) derives a so-called Bayes-Stein estimator from a suitable prior: the shrinkage is a function of the observed dispersion of the sample averages around the grand mean r_0 and of the sample size T:

$$\hat{w} = \frac{\hat{\lambda}}{(T + \hat{\lambda})}$$

$$\hat{\lambda} = \frac{(N + 2)(T - 1)}{(\mathbf{r} - r_0\mathbf{1})' S^{-1}(\bar{\mathbf{r}} - r_0\mathbf{1})(T - N - 2)},$$

where S is the usual sample covariance matrix. With this particular prior, r_0 is shown to be the mean of the minimum variance portfolio. For our set of stock returns, the estimated value of λ is usually around 110; with 5 years of monthly data, the shrinkage factor is 0.65. Finally, for this Bayes-Stein estimator, the covariance matrix can be reasonably approximated by S, if the sample size is moderate.

Armed with a vector of expected returns and a covariance matrix, it is now possible to tackle the analysis of the efficient set. But first recall an important property of the minimum variance frontier.[16] Figure 1

15. For the negative exponential utility function, the optimum weights can be explicitly written as $\mathbf{q} = (1/g)\Sigma^{-1}\boldsymbol{\mu} + [(g - b)/cg]\Sigma^{-1}\mathbf{1}$, with $c = \mathbf{1}'\Sigma^{-1}\mathbf{1}$ and $b = \mathbf{1}'\Sigma^{-1}\boldsymbol{\mu}$, g = constant absolute risk aversion. Replace $\boldsymbol{\mu}$ by $(1 - w)\boldsymbol{\mu} + w\eta\mathbf{1}$. Then b becomes $b^* = (1 - w)b + wc\eta$, and $\mathbf{q}$ becomes, after simplification, $\mathbf{q}^* = (1 - w)\mathbf{q} + w\,(1/c)\Sigma^{-1}\mathbf{1}$, which does not depend on η. The proposition does not hold if the weights do not have to sum to unity, i.e., if there is risk-free lending or borrowing.

16. The minimum variance frontier is defined as the locus of points in the mean-variance plane that minimize the variance of the return on the investment portfolio for a given level of expected return. The efficient set refers to the upward-sloping portion of the minimum variance frontier. This line represents the best attainable combination of risk and expected return.

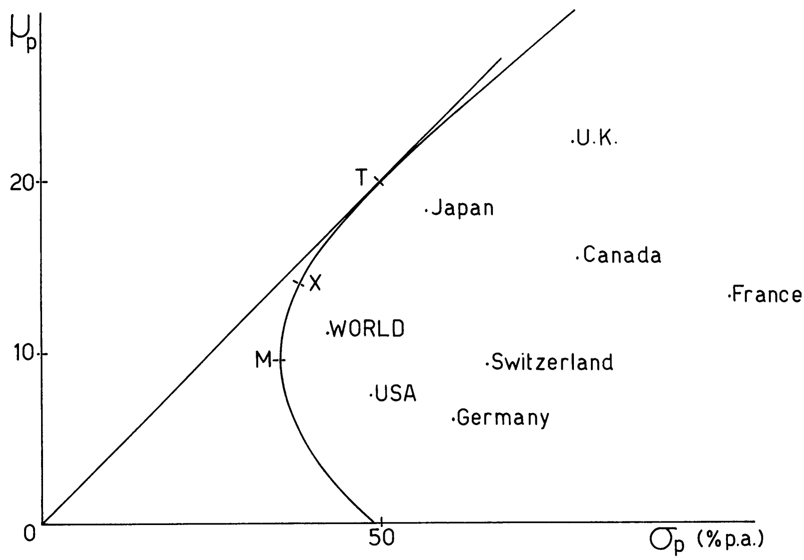

FIG. 1.—Geometry of the efficient set

illustrates a classical minimum variance frontier derived from the data described in table 3. Estimates are based on 5 years of monthly observations ending in December 1981. If short sales are allowed, any portfolio X on the efficient set can be written as a weighted average of two fixed portfolios: the minimum variance portfolio M and the tangent portfolio T with the highest ratio μ/σ. With risk-free lending at a zero rate of interest, this tangent portfolio is the optimal choice for all investors. For the N assets under consideration, the vector of portfolio weights $\mathbf{q}$ can be written as[17]

$$\mathbf{q} = x \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} + (1 - x)\frac{\Sigma^{-1}\boldsymbol{\mu}}{\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu}} = x\mathbf{q}_M + (1 - x)\mathbf{q}_T, \tag{4}$$

where $\boldsymbol{\mu}$ is the vector of expected returns, $\mathbf{1}$ is a vector of ones, and Σ is the variance-covariance matrix of asset returns. In practice, both portfolios M and T have to be derived from sample estimates. The weights of the minimum variance portfolio depend only on the sample covariance matrix; on the other hand, the classical tangent portfolio also relies on sample means. I contend that the latter is quite imprecisely estimated: more emphasis should be placed on the minimum variance portfolio, which is precisely what using Stein estimation amounts to.

Figure 2 depicts the minimum variance frontier for different estimators for expected returns, with the approximation of identical

17. For an analytic derivation of the efficient frontier, see Merton (1972).

TABLE 3 Distribution Parameters for the Efficient Set

	Mean Return	Covariance Matrix						
Canada	1.287	42.18						
France	1.096	20.18	70.89					
Germany	.501	10.88	21.58	25.51				
Japan	1.524	5.30	15.41	9.60	22.33			
Switzerland	.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.A.	.620	17.41	12.62	4.70	1.00	7.20	9.90	16.42
World	.916	12.22						

Other Statistics

$$c = \mathbf{1}'S^{-1}\mathbf{1} = .11838$$
$$b = \mathbf{r}'S^{-1}\mathbf{1} = .09532$$
$$a = \mathbf{r}'S^{-1}\mathbf{r} = .15847$$

NOTE.—Dollar returns in percentage per month. Period covered is January 1977–December 1981.

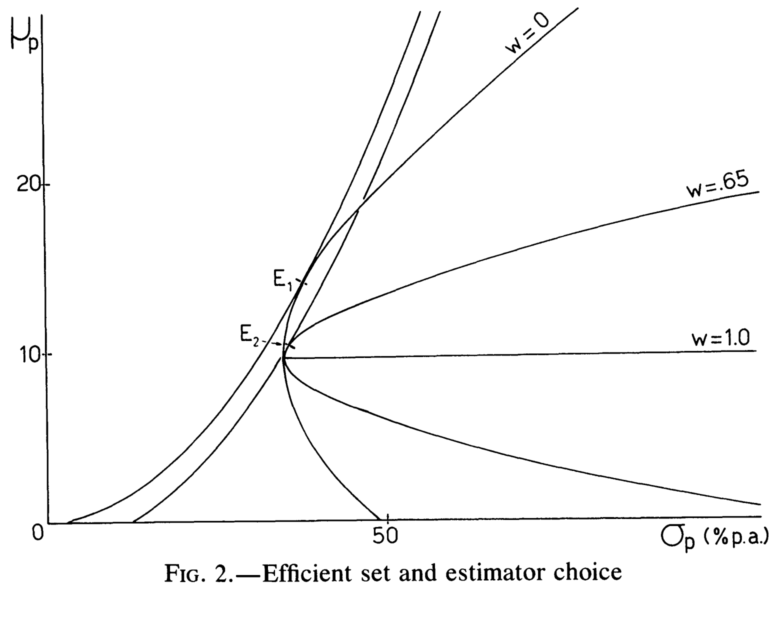

FIG. 2.—Efficient set and estimator choice

covariance matrices. The line indexed by $w = 0$ corresponds to the usual case; for a negative exponential utility function,[18] the optimal portfolio is approximately halfway between the points M and T, the weight on the minimum variance portfolio being 58.6%. With Bayes-

18. This functional form is convenient because it leads to closed-form solutions for the optimum weights. The absolute risk aversion parameter was chosen to be 23 for monthly data. In annual terms, this implies that a 1% increase in the variance (10% in standard deviation) must be accompanied by an increase of about 1% in average return.

Stein estimation and $w = 0.65$, the optimal portfolio comes much closer to the minimum variance point: the proportion x increases to 94.1%. If one accepts the hypothesis that $w = 1$, the best forecast is the common mean, the hyperbola becomes a straight line, and diversification can only lead to risk reduction.

Thus, with increasing shrinkage, the efficient set becomes flatter, and the expected increase in average returns dissipate accordingly. But for the American investor who initially held the U.S. market, the decrease in ex ante variance from going international is still substantial. To check whether this difference is significant, in Appendix B I develop a test statistic for the hypothesis that the variance of one asset, say the U.S. market, is equal to the variance of the minimum variance portfolio. The statistic is based on the likelihood ratio criterion and has an asymptotic χ^2 distribution with $(N - 1)$ degrees of freedom, where N is the number of assets involved. For this data set, the value of the test statistic is 61.6, which implies a very strong rejection of the null hypothesis. Therefore, the risk reduction factor is highly significant: this important result stems from the relative precision of estimating variances.

V. Simulation of Actual Investment Strategies

How practical are these results? I shall now compare three simulated investment strategies: passive, classical, and Bayes-Stein. The classical approach produces results very similar to the Bayes diffuse prior case, whose results are not reported here.

First, a *passive* approach is used as a benchmark. The investor could hold the U.S. index, the world index, or an equally weighted index. Second, examine the *classical* strategy. As Grubel (1968), Levy and Sarnat (1970), and more recently Solnik (1982) and Adler and Dumas (1983) have suggested, the investor chooses a portfolio on the efficient frontier, which is constructed on the basis of past sample estimates at each point in time. With risk-free lending at a zero rate of interest,[19] the investor picks the portfolio with the highest ratio of expected return to standard deviation, which lies on the tangent from the origin. Another investment strategy is to choose an optimal portfolio on the basis of a fixed utility function, say the negative exponential.

Third, the *Bayes-Stein* approach would shrink these past averages toward a common value, with optimal shrinkage derived from the data. If one accepts the extreme view that all expected returns are equal,

19. With a positive risk-free rate of interest, the optimal portfolio will be even higher than the tangent portfolio, and any undesirable characteristic of the tangent portfolio will be accentuated. Therefore, my approach leads to a conservative measure of the effect of estimation risk.

then one should invest only in the minimum variance portfolio, regardless of the information in the sample means.

The investment strategies considered were implemented on the actual data set, and the results are displayed in table 4. Adaptive portfolios were formed each period on the basis of observed returns over the last 60 months. From the subsequent time series of realized returns, I derived ex post means and variances; averages of ex ante values are also reported. Ex post performances are compared on the basis of the ratio μ/σ. Using the methodology proposed by Jobson and Korkie (1981), I tested the hypothesis that the ratios were equal.[20]

Perhaps the most striking result of table 4 is the discrepancy between ex ante and ex post means, another indication of the poor forecasting ability of past averages. Standard deviations, on the other hand, are closer to their ex ante values, with a generally upward bias. As a result, the tangent portfolio, constructed so as to maximize the ratio μ/σ, is the worst performer ex post: this strategy succeeds in increasing the variance but not the average return of the portfolio. The performance of the utility-derived portfolios is somewhat better, because they are closer to the minimum variance portfolio.

In contrast, the Bayes-Stein approach leads to improved results for all portfolios. The minimum variance portfolio also performs quite well: the ex post standard deviation is the lowest, as it was supposed to be, but also the mean is comparatively high for this time period. This is an astonishing performance, which seems to imply that 5-year averages convey absolutely no information about the future. The simulation analysis in Jorion (1984) confirms this result for sample sizes up to 60; however, for larger samples, the Bayes-Stein strategy improved both on the classical and the minimum variance portfolio. Unfortunately, increasing the sample size is difficult here because the total number of observations is limited to 13 years.

It remains to be seen whether these results could be due to chance. The second panel of table 4 reports pairwise tests of equal performance. Given the reported low power of the test statistic,[21] it is remarkable to find two significant entries: the tangent portfolio is significantly outperformed by the Bayes-Stein and minimum variance

20. The tests proposed by Jobson and Korkie (1981) are based on stationary excess returns rather than nominal returns. Nominal returns are used frequently in international finance, and it is very difficult to discriminate between the two models on the basis of the data. The test statistic is

$$t_{ij} = \frac{s_j \bar{r}_i - s_i \bar{r}_j}{[2/T(s_i^2 s_j^2 - s_i s_j s_{ij})]^{1/2}} .$$

21. Jobson and Korkie (1981) report that for a typical experiment ($\mu_1 = 1.2$, $\mu_2 = 0.6$, $\sigma_1 = 4$, $\sigma_2 = 3$, $\rho = 0.5$) the power of the test was only 15% for a type I error level of 5% and sample sizes up to 150. Thus, even if the null were false, it would not be rejected about 85% of the time.

TABLE 4 **Comparisons of Ex Ante and Ex Post Efficient Sets**

A. Expected Returns Based on 60-Month Averages
(January 1976–December 1983)

Strategy	Mean: Ex Ante (Average)	Mean: Ex Post	SD: Ex Ante (Average)	SD: Ex Post	Mean/SD (Ex Post)
Passive:					
U.S.A.	.554	.977	4.600	4.039	.242
Equal weights	.800	.955	4.626	4.048	.236
World	.696	1.014	4.105	3.524	.288
Classical:					
Tangent	1.873	.299	7.300	12.202	.024
Negative exponential utility	1.006	1.088	3.647	3.933	.277
Bayes-Stein:					
Tangent	.849	.947	4.225	4.665	.203
Negative exponential utility	.791	1.353	3.813	4.051	.334
Minimum variance	.779	1.219	3.509	3.571	.341

B. Pairwise Tests of Equal Performance

	Portfolio: World	Portfolio: Classical Tangent	Portfolio: Bayes-Stein Tangent
Classical	1.93		
Bayes-Stein	.88	−3.08**	
Minimum variance	−1.01	−2.44*	−1.72

Note.—Returns in percentage per month. At each point in time, the portfolio is computed on the basis of the last 60 observations for the seven countries. The first such ex ante portfolio is based on data from January 71 to December 75. Ex post data available from January 76 to December 83 (T = 96). Tests of ex post performance are based on the tests of equality of Sharpe ratios developed by Jobson and Korkie (1981).

* Significant at the 5% level.

** Significant at the 1% level.

portfolios. Other comparisons are not so conclusive. In particular, the best strategy, holding the minimum variance portfolio, is not significantly superior to simply holding the world index.

The comparative performance of the active portfolios can be better understood by examining the behavior of the optimum weights. In figure 3 the proportion of the U.S. weight in two adaptive portfolios is plotted against time. The minimum variance portfolio relies solely on variances and covariances; since these are rather precisely estimated, the composition of the portfolio is quite stable over time. For the tangent portfolio, on the other hand, the weight of the U.S. index fluctuates wildly. In some periods, the tangent portfolio takes extreme positions, resulting in wide variations in ex post returns. These erratic fluctuations explain why the classical portfolio does so poorly outside the sample period.

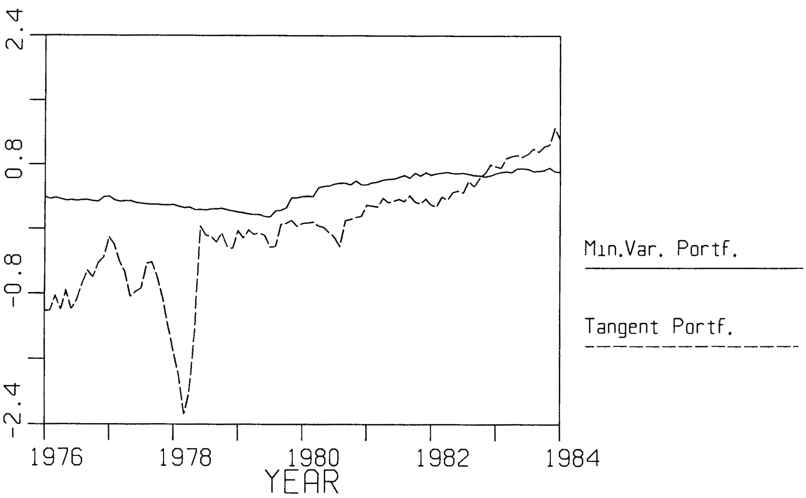

FIG. 3.—Weight of United States in optimal portfolio

In table 5 I also compare the composition of typical ex post optimal portfolios based on the 1977–81 data analyzed in the previous section. High weights are assigned to countries with high past means; the U.S. weight in the tangent portfolio is only 23%. On the other hand, this proportion is 65% in the minimum variance portfolio: this is due to the low variance of the returns on the U.S. index, partly explained by the size of the market and the absence of exchange rate movement. Most of the remaining weight is placed on Germany and Japan.

For illustrative purposes, I also computed optimum weights if all asset prices were measured in French francs. The tangent portfolio is similar to the U.S. tangent portfolio because the relative ranking of past means is preserved.[22] The covariance matrix differs, however, when assets prices are measured in different currencies. This occurs because exchange rate changes are impounded on the variance of foreign equity returns. This effect is marked on the minimum variance portfolio, where the proportion assigned to the United States decreases from 65% to 40%. The weight assigned to the French market increases, but not by much; this small value can be explained by the high variability of the returns on the French index.

VI. Conclusions

International portfolio diversification has too often been analyzed in the classical mean-variance framework where no attention is paid to

22. If stock prices and exchange rates can be modeled by diffusion processes, the portfolios should be identical, as demonstrated by Sercu (1980).

TABLE 5 Optimum Portfolio Weights

	Canada	France	Germany	Japan	Switzerland	U.K.	U.S.A.
	Market Weights						
	.05	.03	.06	.18	.03	.10	.55
	Returns Measured in U.S. Dollars						
Portfolio:							
Tangent	.05	−.08	−.22	.69	.01	.32	.23
Minimum variance	−.13	−.11	.30	.36	−.11	.04	.65
Optimum:							
Negative exponential utility:							
$w = 0$	−.05	−.10	.08	.50	−.06	.15	.48
$w = .65$	−.10	−.10	.22	.40	−.09	.08	.59
	Returns Measured in French Francs						
Portfolio:							
Tangent	.04	−.10	.01	.70	−.07	.28	.14
Minimum variance	−.06	−.09	.32	.47	−.13	.09	.40

NOTE.—Ex post optimum weights. Estimation period is January 1977–December 1981.

uncertainty about the expected value and covariance matrix of asset returns. This paper argues that estimation risk due to uncertain mean returns has a considerable impact on optimal portfolio selection and that alternative estimators for expected returns should be explored. In contrast, variances and covariances are measured with relative precision. Among the shortcomings of the classical approach, most notable are the instability of portfolio weights and the sharp deterioration of performance when out-of-sample data are used.

Next I discuss an alternative estimator for expected returns, which discounts the influence of extreme sample means: the Stein estimator. The essential implication of this estimator is that portfolio selection should rely more heavily on the minimum variance portfolio, which is the optimal choice if all expected returns were equal. An active portfolio allocation experiment indicates that the classical approach is systematically outperformed by a strategy based on these so-called shrinkage estimators.

In summary, in this paper I highlighted the pitfalls of analyzing portfolio diversification in a mean-variance framework based on ex post data: I show that the usual conclusions vastly overestimate the extent of possible gains in average returns. Instead I propose a more conservative estimator for expected returns, which suggests that most of diversification benefits are likely to accrue from a reduction in risk.

Appendix A

Data Sources and Definition

National equity indices were compiled from the publication *Capital International Perspective*, starting in January 1971. This source provides monthly value-weighted stock market indices in domestic currency, dividend yields, and exchange rates against the U.S. dollar. Dividends are reported as last year's dividends divided by 12. I converted the domestic indices into dollar indices, with dividends reinvested. Returns are defined in logarithmic form throughout. Summary statistics for the data can be found in Ibbotson, Carr, and Robinson (1982).

Appendix B

Test of Risk Reduction Hypothesis

The hypothesis of interest in this situation is whether a significant reduction in risk can be achieved through diversification. Formally, I wish to test whether the variance of the minimum variance portfolio σ_0^2 is equal to the variance of one of the portfolio assets σ_1^2, chosen to be the first one without loss of generality. Since the two variances are not independent, the standard F-test cannot be used.

First expand the expression for the variance of the minimum variance portfolio $\sigma_0^2 = 1/(\mathbf{1}'\Sigma^{-1}\mathbf{1})$. As before, boldface letters represent vectors. Partition the matrix Σ as

$$\Sigma = \begin{vmatrix} \sigma_{11} & \boldsymbol{\sigma}_{12}' \\ \boldsymbol{\sigma}_{12} & \Sigma_{22} \end{vmatrix}.$$

The partitioned inverse Σ^{-1} can be written as

$$\begin{vmatrix} \sigma_{11}^{-1} + \sigma_{11}^{-1}\boldsymbol{\sigma}_{12}'\Sigma_{22\cdot1}^{-1}\boldsymbol{\sigma}_{12}\sigma_{11}^{-1} & -\sigma_{11}^{-1}\boldsymbol{\sigma}_{12}'\Sigma_{22\cdot1}^{-1} \\ -\Sigma_{22\cdot1}^{-1}\boldsymbol{\sigma}_{12}\sigma_{11}^{-1} & \Sigma_{22\cdot1}^{-1} \end{vmatrix},$$

where $\Sigma_{22\cdot1}$ is defined as $\Sigma_{22} - \boldsymbol{\sigma}_{12}(1/\sigma_{11})\boldsymbol{\sigma}_{12}'$. We have thus

$$\mathbf{1}'\Sigma^{-1}\mathbf{1} = (1/\sigma_{11}) + (\mathbf{1} - \boldsymbol{\sigma}_{12}/\sigma_{11})'\Sigma_{22\cdot1}^{-1}(\mathbf{1} - \boldsymbol{\sigma}_{12}/\sigma_{11}),$$

and, under H_0, $\mathbf{1}'\Sigma^{-1}\mathbf{1} = (1/\sigma_{11})$, the second term on the right-hand side of the above equation must be zero. Since the matrix $\Sigma_{22\cdot1}$ is positive definite if Σ is positive definite, the quadratic form will be zero only if $(\mathbf{1} - \boldsymbol{\sigma}_{12}/\sigma_{11}) = \mathbf{0}$. The hypothesis H_0 can therefore be written as H_0: $\boldsymbol{\sigma}_{12} = \mathbf{1}\,\sigma_{11}$.

A test procedure for H_0 can be developed using the likelihood ratio principle. The logarithm of the likelihood function for T independent samples $\mathbf{x}$ from a multivariate normal distribution $N(\boldsymbol{\mu}, \Sigma)$ is given by

$$\ln \mathrm{L} = -(NT/2)\ln 2\pi - (T/2)\ln|\Sigma| - (1/2)\sum_{t=1}^{T}(\mathbf{x}_t - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}_t - \boldsymbol{\mu}),$$

where N is the size of the vector $\mathbf{x}$. The test statistic is the ratio of the likelihood function maximized under the null hypothesis over the likelihood function maximized with no restrictions. Without restrictions, the logarithm of the likelihood function is maximized at

$$\ln L^* = -(NT/2)\ln 2\pi - (T/2)\ln|\hat{\Sigma}| - (NT/2),$$

for

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} \equiv \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t,$$

$$\hat{\Sigma} = S \equiv \frac{1}{T}\sum_{t=1}^{T}(\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})'.$$

In order to test H_0, it is necessary to maximize the likelihood function under the restriction $\boldsymbol{\sigma}_{12} = \mathbf{1}\sigma_{11}$. Under H_0, Σ becomes Σ_0, and

$$\ln L_0 = -(NT/2)\ln 2\pi - (T/2)\ln|\Sigma_0| - (1/2)\sum_{t=1}^{T}(\mathbf{x}_t - \bar{\mathbf{x}})'\Sigma_0^{-1}(\mathbf{x}_t - \bar{\mathbf{x}})$$
$$-(1/2)T(\bar{\mathbf{x}} - \boldsymbol{\mu})'\Sigma_0^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}).$$

Since H_0 imposes no restriction on the vector $\boldsymbol{\mu}$, the maximum likelihood estimator of $\boldsymbol{\mu}$ is the same as before, and the last term drops when evaluating the maximum of $\ln L_0$ over $\boldsymbol{\mu}$. Partition Σ_0 as above, and write the determinant $|\Sigma|$ as $\sigma_{11}|\Sigma_{22\cdot 1}|$. This gives

$$\begin{aligned}\ln L_0 = {} & -(NT/2)\ln 2\pi - (T/2)\ln\sigma_{11} - (T/2)\ln|\Sigma_{22\cdot 1}| \\ & -(1/2)\sum_{t=1}^{T}(x_{1t} - \bar{x}_1)'(1/\sigma_{11})(x_{1t} - \bar{x}_1) \\ & -(1/2)\sum_{t=1}^{T}[(\mathbf{x}_{2t} - \bar{\mathbf{x}}_2) - (x_{1t} - \bar{x}_1)\mathbf{1}]'\Sigma_{22\cdot 1}^{-1}[(\mathbf{x}_{2t} - \bar{\mathbf{x}}_2) - (x_{1t} - \bar{x}_1)\mathbf{1}],\end{aligned}$$

where the last term was obtained using the restriction H_0. The likelihood function can therefore be separated into two components, the marginal likelihood for x_1 and the conditional likelihood for $\mathbf{x}_2$, given x_1. The maximum likelihood estimators under H_0 are

$$\hat{\sigma}_{11} = \frac{1}{T}\sum_{t=1}^{T}(x_{1t} - \bar{x}_1)^2,$$

$$\begin{aligned}\hat{\Sigma}_{22\cdot 1} &= \frac{1}{T}\sum_{t=1}^{T}[(\mathbf{x}_{2t} - \bar{\mathbf{x}}_2) - (x_{1t} - \bar{x}_1)\mathbf{1}][(\mathbf{x}_{2t} - \bar{\mathbf{x}}_2) - (x_{1t} - \bar{x}_1)\mathbf{1}]' \\ &= \hat{\Sigma}_{22} - 2\,\hat{\boldsymbol{\sigma}}_{12}\,\mathbf{1}' + \hat{\sigma}_{11}\,\mathbf{1}\,\mathbf{1}',\end{aligned}$$

and the logarithm of the maximized likelihood function is

$$\begin{aligned}\ln L_0^* = {} & -(NT/2)\ln 2\pi - (T/2)\ln[\hat{\sigma}_{11}|\hat{\Sigma}_{22} \\ & - 2\,\hat{\sigma}_{12}\mathbf{1}' + \hat{\sigma}_{11}\mathbf{1}\,\mathbf{1}'|] - (NT/2).\end{aligned}$$

Therefore, the log-likelihood ratio can be written as

$$\ln\lambda = -(T/2)\ln[\hat{\sigma}_{11}|\hat{\Sigma}_{22} - 2\hat{\boldsymbol{\sigma}}_{12}\mathbf{1}' + \hat{\sigma}_{11}\mathbf{1}\,\mathbf{1}'|] + (T/2)\ln|\hat{\Sigma}|$$

and the test statistic

$$-2 \ln \lambda = T \ln \frac{\hat{\sigma}_{11}|\hat{\Sigma}_{22} - 2\hat{\boldsymbol{\sigma}}_{12}\mathbf{1}' + \hat{\sigma}_{11}\mathbf{1}\,\mathbf{1}'|}{|\hat{\Sigma}|}$$

has under the null a χ^2 distribution with $(N - 1)$ degrees of freedom.

References

Adler, M., and Dumas, B. 1983. International portfolio choice and corporation finance: A survey. *Journal of Finance* 38 (June): 925–84.

Brown, S. J. 1979. The effect of estimation risk on capital market equilibrium. *Journal of Financial and Quantitative Analysis* 14 (June): 215–20.

Efron, B., and Morris, C. 1973. Stein's estimation rule and its competitors: An empirical Bayes approach. *Journal of the American Statistical Association* 68 (March): 117–30.

Efron, B., and Morris, C. 1975. Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association* 70 (June): 311–19.

Fama, E. F. 1976. Forward rates as predictors of future spot rates. *Journal of Financial Economics* 3 (October): 361–78.

Fama, E. F., and Farber, A. 1979. Money, bonds and foreign exchange. *American Economic Review* 69 (September): 639–49.

Grauer, F. L.; Litzenberger, R. H.; and Stehle, R. E. 1976. Sharing rules and equilibrium in an international capital market under uncertainty. *Journal of Financial Economics* 3 (June): 233–56.

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

Ibbotson, R. G.; Carr, R. C.; and Robinson, A. W. 1982. International equity and bond returns. *Financial Analysts Journal* 38 (July-August): 61–83.

James, W., and Stein, C. 1961. "Estimation with Quadratic Loss." *Proceedings of the 4th Berkeley Symposium on Probability and Statistics* 1. Berkeley: University of California Press.

Jobson, J. D.; Korkie, B.; and Ratti, V. 1979. "Improved Estimation for Markowitz Portfolios Using James-Stein Type Estimators." *Proceedings of the American Statistical Association, Business and Economics Statistics Section*. Washington: American Statistical Association.

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

Jorion, P. 1984. Bayes-Stein estimation for portfolio analysis. Working paper. New York: Columbia University.

Klein, R. W., and Bawa, V. S. 1976. The effect of estimation risk on optimal portfolio choice. *Journal of Finance Economics* 3 (June): 215–31.

Levy, H. 1981. Optimal portfolio of foreign currencies with borrowing and lending. *Journal of Money, Credit and Banking* 13 (August): 325–41.

Levy, H., and Sarnat, M. 1970. International diversification of investment portfolios. *American Economic Review* 60 (September): 668–75.

Levy, H., and Sarnat, M. 1975. Devaluation risk and the portfolio analysis of international investments. In E. Elton and M. Gruber, eds., *International Capital Markets*. Amsterdam: North-Holland.

Levy, H., and Sarnat, M. 1979*a*. Exchange risk and the optimal diversification of foreign currency holdings. *Journal of Money, Credit and Banking* 10 (November): 453–63.

Levy, H., and Sarnat, M. 1979*b*. Foreign exchange risk, portfolio diversification, and the demand for money. In M. Sarnat and G. Szego, eds., *International Finance and Trade*. Cambridge, Mass.: Ballinger.

Lindley, D. V. 1962. Discussion on Professor Stein's paper. *Journal of Royal Statistical Society* 24B, pp. 285–87.

Logue, D. E. 1982. An experiment in international diversification. *Journal of Portfolio Management* 9 (Fall): 22–27.

Losq, E. 1979. Diversification and asset valuation in an international capital market. In

M. Sarnat and G. Szego, eds., *International Trade and Finance*. Cambridge, Mass.: Ballinger.

Markowitz, H. M. 1959. *Portfolio Selection: Efficient Diversification of Investments*. New York: Wiley.

Merton, R. E. 1972. An analytic derivation of the efficient portfolio frontier. *Journal of Financial and Quantitative Analysis* 7 (September): 1851–72.

Merton, R. C. 1980. On estimating the expected return on the market. *Journal of Financial Economics* 8 (December): 323–61.

Morris, C. N. 1983. Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association* 78 (March): 47–55.

Sercu, P. 1980. A generalization of the international asset pricing model. *Revue de l'Association Française de Finance* 1 (June): 91–135.

Solnik, B. 1974*a*. An equilibrium model of the international capital market. *Journal of Economic Theory* 8 (August): 500–524.

Solnik, B. 1974*b*. The international pricing of risk: An empirical examination of the world capital market structure. *Journal of Finance* 29 (May): 365–78.

Solnik, B. 1977. Testing international asset pricing: Some pessimistic views. *Journal of Finance* 32 (May): 503–12.

Solnik, B. 1982. Optimal international asset allocation. *Journal of Portfolio Management* 9 (Fall): 11–21.

Stehle, R. E. 1977. An empirical test of the alternative hypotheses of national and international pricing of risky assets. *Journal of Finance* 32 (May): 493–502.

Stein, C. 1955. "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the 3d Berkeley Symposium on Probability and Statistics* 1. Berkeley: University of California Press.

Stultz, R. 1981. A model of international asset pricing. *Journal of Financial Economics* 9 (December): 384–406.

Zellner, A., and Chetty, V. K. 1965. Prediction and decision problems in regression models from the Bayesian point of view. *Journal of the American Statistical Association* 60 (June): 608–15.

Zellner, A., and Vandaele, W. 1974. Bayes-Stein estimators for *k*-means regression and simultaneous equations models. In S. Fienberg and A. Zellner, eds., *Studies in Bayesian Econometrics and Statistics*. Amsterdam: North-Holland.