NBER WORKING PAPER SERIES

ANALYZING INVESTMENTS WHOSE HISTORIES DIFFER IN LENGTH

Robert F. Stambaugh

Working Paper 5918

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 1997

I am grateful to Kerry Back, Gordon Bodnar, Francisco Delgado, Shmuel Kandel, Lubos Pastor, G. William Schwert, Ross Stevens, workshop participants at Duke University, the University of North Carolina, the University of Pennsylvania, the University of Rochester, Washington University in St. Louis, Yale University, and especially Jay Shanken (the referee) for helpful comments and discussions. This paper is part of NBER's research program in Asset Pricing. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

Analyzing Investments Whose Histories Differ
in Length
Robert F. Stambaugh
NBER Working Paper No. 5918
February 1997
JEL Nos. G11, G15, C11
Asset Pricing

ABSTRACT

This study explores multivariate methods for investment analysis based on a sample of return histories that differ in length across assets. The longer histories provide greater information about moments of returns, not only for the longer-history assets, but for the shorter-history assets as well. To account for the remaining parameter uncertainty, or "estimation risk," portfolio opportunities are characterized by a Bayesian predictive distribution. Examples involving emerging markets demonstrate the value of using the combined sample of histories and accounting for estimation risk, as compared to truncating the sample to produce equal-length histories or ignoring estimation risk by using maximum-likelihood estimates.

Robert F. Stambaugh
Finance Department
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104-6367
and NBER
stambaugh@wharton.upenn.edu

1. Introduction

Historical rates of return are often used in investment analysis. Estimates of moments of returns based on historical time series provide information useful in selecting portfolios, evaluating investment performance, and investigating models of asset pricing. In many applications, the lengths of available histories differ across the assets being analyzed, especially when the assets are traded on separate exchanges or in different countries. For example, substantial differences in lengths of histories are likely to occur, virtually by definition, when the universe of assets includes investments in emerging markets. A typical approach to this problem, especially in applications of multivariate methods, is to base the historical analysis on a sample in which all return histories begin at a common date. That is, the longer return histories of the "developed-market" assets are truncated, so that any returns observed before the available history of the emerging-market investments are simply discarded.[1]

In many cases, it is neither necessary nor desirable to discard returns. Suppose, for example, that the researcher or decision maker would use some of those discarded returns if the shorter-history assets were not included in the analysis.[2] Then, in general, those discarded returns contain information that is useful in an analysis that includes the shorter-history assets. Not only do those discarded returns provide additional information about the longer-history assets, but they generally provide information about the shorter-history assets as well.

This study investigates multivariate methods that use a "combined" sample in which the lengths of return histories differ across assets. Although such methods could be developed under a variety of assumed probability distributions for returns, the i.i.d. multivariate Normal model assumed here permits closed-form analytic results that simplify the essential ideas. Moreover, that assumption is often employed in studies that propose multivariate methods for samples of equal-length return histories.[3] It is hoped that the results obtained here in the standard setting motivate extensions to richer stochastic frameworks.

[1] See Harvey (1995) for a recent example.

[2] In other words, those returns would not be discarded due to a concern that the stochastic framework assumed for the longer-history assets does not hold for any period longer than that used in the truncated sample. One might note that such a concern is not evident in previously published empirical work: empirical studies that do not include emerging markets, for example, seldom if ever choose a first sample date that happens to coincide with the beginning of an emerging-markets data set.

[3] Examples include the likelihood ratio test of a portfolio's mean-variance efficiency in Gibbons, Ross, and Shanken (1989) and the Bayesian analysis of a portfolio's degree of inefficiency in Kandel, McCulloch, and Stambaugh (1995).

The paper is organized as follows. Maximum-likelihood estimates (MLE's) of first and second moments are presented in section 2, and the combined-sample MLE's are compared to the more common truncated-sample estimates. When the parameters of the return distribution must be estimated from a finite sample of returns, then the imprecision in those estimates presents an investor with additional uncertainty, or "estimation risk." This estimation risk is reflected in the Bayesian predictive distribution of future returns. Section 3 derives the first and second moments of that predictive distribution, conditioned on the combined sample. In order to focus on the essential concepts, the analyses in sections 2 and 3 are limited to the case where each asset's history begins at one of only two possible dates. In practice, starting dates are often more heterogeneous, and section 4 extends the results in sections 2 and 3 to an arbitrary number of different starting dates.

Sections 5 and 6 illustrate the empirical methods using monthly data in portfolio problems involving emerging markets. Section 5 analyzes a mean-variance optimization problem with an asset universe consisting of one-month U.S. Treasury bills (assumed riskless) and three risky index portfolios: (i) Standard & Poor's composite index, (ii) Morgan Stanley Capital International's index for Europe, Australia, and the Far East, and (iii) the International Finance Corporation's (IFC) composite index for emerging markets. Returns beginning in 1970 are used for the first two indices, whereas the emerging-market returns begin in 1985. An optimal portfolio constructed using the combined sample and accounting for estimation risk can be compared to a portfolio that is constructed using only the post-1985 data. If the latter construction ignores estimation risk as well, then an investor with relative risk aversion equal to 3 would value that suboptimal portfolio less than the optimal portfolio by about 23 basis points per month, in terms of certainty-equivalent return.

Section 6 considers the problem of constructing the minimum-variance portfolio from a universe of 22 emerging-market index portfolios. Each index portfolio is designed by the IFC to reflect the portion of a given country's equity market that is accessible to foreign investors. For the 22 countries included in this example, the first month of available data ranges from January 1989 to November 1993. To an investor who uses the combined sample of all available histories and accounts for estimation risk, the minimum-variance portfolio has a standard deviation of about 3.8 percent per month. That same investor assigns a standard deviation of at least 6.1 percent to portfolios constructed using methods that either ignore estimation risk or discard returns on the longer-history assets.

The examples in sections 5 and 6 illustrate the conditional Bayesian decision approach, wherein the investor bases decisions on the predictive distribution that is conditioned on the

single observed sample. As those examples demonstrate, the predictive distribution can also be used to assess the costs associated with various suboptimal choices, such as portfolios formed by methods that truncate the sample or ignore estimation risk. An alternative "frequentist" approach to evaluating the relative merits of various portfolio selection methods is to compare their performances in repeated hypothetical random samples, where performance is evaluated using "true" moments of returns. Section 7 conducts such investigations in settings similar to those of the examples in sections 5 and 6. The results confirm the potentially substantial costs associated with truncating the sample or ignoring estimation risk. Section 8 concludes the study with a brief discussion of possible extensions.

2. Maximum-Likelihood Estimation

2.1. Stochastic Setting

Let the vector $R_{1,t}$ contain the returns on N_1 assets in period t, and assume there are T observations of these returns for periods $1, \ldots, T$. The T observations of $R_{1,t}$ are assumed to be independent realizations from a multivariate normal distribution with

$$\mathrm{E}\{R_{1,t}\} = E_1 \tag{1}$$

and

$$\mathrm{cov}\{R_{1,t}, R_{1,t}'\} = V_{11}. \tag{2}$$

Let the vector $R_{2,t}$ contain the returns on another set of N_2 assets in period t, and assume that these returns are observed only for periods $s, \ldots, T$, where $s \geq 1$. For any period $t \geq s$, let $R_t = [R_{1,t}' \; R_{2,t}']'$ denote the combined vector of $N = N_1 + N_2$ returns, and let $S = T - s + 1$ denote the number of observation of this combined vector of returns. It is assumed that, given the starting period s, the S observations of R_t are independent realizations from a multivariate normal distribution with

$$\mathrm{E}\left\{\begin{bmatrix} R_{1,t} \\ R_{2,t} \end{bmatrix}\right\} = \begin{bmatrix} E_1 \\ E_2 \end{bmatrix} = E \tag{3}$$

and

$$\mathrm{cov}\left\{\begin{bmatrix} R_{1,t} \\ R_{2,t} \end{bmatrix}, \begin{bmatrix} R_{1,t}' & R_{2,t}' \end{bmatrix}\right\} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} = V, \tag{4}$$

where V is nonsingular. It is also assumed that $S > \max(N_1, N_2 + 2)$. When, as in the examples presented later, the historical data are used for investment decisions, the above distribution is also assumed for the N returns in period $T + 1$.

The starting period s for the short-history assets is assumed to be either non-stochastic or drawn from a distribution that does not depend on E or V, conditional on the long-history-asset returns $R_{1,t}$, $t = 1, \ldots, T$. This assumption permits some randomness and endogeneity in the starting period, provided that s does not contain information about E and V beyond that contained in the sample of observed returns. This assumption is satisfied, for example, in the simple case where a function of the observed sample history of $R_{1,t}$ determines s: high realized returns on existing assets might give rise to more assets.[4] If, however, s also depends on realizations of returns on the *short*-history assets prior to s, returns not included in the sample, then the assumption made here is generally not satisfied.[5]

In some applications, it may be that the second set of N_2 assets existed before period s but their returns are not included in the sample. In other cases, it may be that those assets did not even exist before period s. In any case, it is not assumed that those assets' actual returns, if the assets did exist, or their hypothetical returns, if they did not exist, obey the same joint distribution for all N assets assumed after period s. If the moments of all N returns prior to period s are given by (3) and (4), but with E_2, V_{12}, and V_{22} replaced by additional free parameters, such a change does not affect the likelihood function for E and V based on the sample of returns used in the analysis. In fact, one might exclude the returns on the N_2 short-history assets before period s, even when such returns are available, because one suspects they are *not* drawn from the same joint distribution as those beginning in period s.[6] What *is* assumed regarding stationarity of distributions is that the marginal distribution of returns on the N_1 long-history assets is the same across all T periods.

One final point to be emphasized about the stochastic setting is that the inclusion of any asset in the sample is, by construction, conditioned on the survival of that asset through period T. If there exist periods during an asset's history in which there were non-zero probabilities that the asset would disappear, then the asset's sample moments include "survival" effects, which generally increase with the probability of disappearance.[7] The methods explored in this study do not incorporate survival probabilities. In that sense, the normality assumption must be viewed as characterizing an asset's return distribution for any period t, conditional on the asset's surviving that period.[8] Of course, relying on such a conditional

[4] I am grateful to Jay Shanken for suggesting this example.

[5] Goetzmann and Jorion (1996) argue that the IFC's decision to add a country to its list of emerging markets depends on previous returns on that country's stock market. This type of endogeneity in the starting dates of short-history assets would violate the assumption made here.

[6] I am grateful to Ross Stevens, of Goldman Sachs Asset Management (GSAM), for suggesting this possibility. The Quantitative Research Group at GSAM has applied this paper's methods in such cases.

[7] See, for example, Brown, Goetzmann, and Ross (1995).

[8] Note that such a result does not obtain, for example, if the return's unconditional distribution is normal

distribution for inference or decision making without also incorporating the probability of disappearance could be unwise, especially if that probability is substantial. An interesting direction for future research would be to extend the methods presented here to incorporate survival effects and more general endogeneity in the starting and stopping times of asset histories.

2.2. Likelihood Function

Define the $S \times N$ matrix

$$Y_S = \left[\begin{array}{cc} Y_{1,S} & Y_{2,S} \end{array} \right] = \left[\begin{array}{cc} R'_{1,s} & R'_{2,s} \\ R'_{1,s+1} & R'_{2,s+1} \\ \vdots & \vdots \\ R'_{1,T} & R'_{2,T} \end{array} \right] \tag{5}$$

and the $T \times N_1$ matrix

$$Y_{1,T} = \left[\begin{array}{c} R'_{1,1} \\ R'_{1,2} \\ \vdots \\ R'_{1,T} \end{array} \right]. \tag{6}$$

The joint density for $Y_{1,T}$ and $Y_{2,S}$, given E, V, and s, follows directly from the assumed multivariate normality and independence across periods:

$$\begin{aligned} p(Y_{1,T}, Y_{2,S}|E,V,s) &= \prod_{t=1}^{s-1} \left(\frac{1}{(2\pi)^{N_1/2}} |V_{11}|^{-1/2} \exp\left\{ -\frac{1}{2}(R_{1,t} - E_1)' V_{11}^{-1} (R_{1,t} - E_1) \right\} \right) \\ &\quad \times \prod_{t=s}^{T} \left(\frac{1}{(2\pi)^{N/2}} |V|^{-1/2} \exp\left\{ -\frac{1}{2}(R_t - E)' V^{-1} (R_t - E) \right\} \right). \end{aligned} \tag{7}$$

When viewed as a function of the parameters, given s and the observed returns, (7) is the likelihood function for E and V.

To see that the likelihood function in (7) obtains even when s is stochastic in the sense described earlier, let ζ denote a vector of parameters, in addition to E and V, that enter the joint density for returns and s. Then the latter joint density can be written

$$\begin{aligned} p(Y_{1,T}, Y_{2,S}, s|E,V,\zeta) &= p(Y_{1,T}, s|E,V,\zeta) \cdot p(Y_{2,S}|Y_{1,T}, s, E, V, \zeta) \\ &= p(s|Y_{1,T}, E, V, \zeta) \cdot p(Y_{1,T}|E,V,\zeta) \cdot p(Y_{2,S}|Y_{1,T}, s, E, V, \zeta) \\ &= p(s|Y_{1,T}, \zeta) \cdot p(Y_{1,T}, Y_{2,S}|E,V,s), \end{aligned} \tag{8}$$

and survival is determined by a minimum-return threshold.

where the last equality follows from the assumed normal distributions of returns, which do not depend on ζ, and the assumption that, conditional on $Y_{1,T}$, the distribution of s does not depend on E or V. The likelihood function for E, V, and ζ—the joint density in (8) viewed as a function of those parameters given the sample—involves E and V only in the second factor, which is the joint density given in (7). The proportionality constant, $p(s|Y_{1,T},\zeta)$, plays no role in obtaining the maximum-likelihood estimators of E and V.

2.3. Estimators

A common approach to estimating E and V is to compute the "truncated-sample" maximum-likelihood estimators based on the S periods in which returns on all N assets are observed. These truncated-sample MLE's of E and V are given by

$$\hat{E}_S = \begin{bmatrix} \hat{E}_{1,S} \\ \hat{E}_{2,S} \end{bmatrix} = \frac{1}{S} Y_S' \iota_S \tag{9}$$

and

$$\hat{V}_S = \begin{bmatrix} \hat{V}_{11,S} & \hat{V}_{12,S} \\ \hat{V}_{21,S} & \hat{V}_{22,S} \end{bmatrix} = \frac{1}{S}(Y_S - \iota_S \hat{E}_S')'(Y_S - \iota_S \hat{E}_S'), \tag{10}$$

where ι_S denotes an S-vector of ones and the partitioning in (9) and (10) follows that of Y_S in (5).

The above truncated-sample estimators do not use the first $s-1$ observations of $R_{1,t}$, which appear in the first factor in (7). Maximizing (7) with respect to the elements of E and V is complicated by the fact that E_1 and V_{11} appear by themselves in the first factor but as submatrices of E and V in the second factor. Following Anderson (1957), however, an analytic solution to the maximization is obtained by performing a change of variables and rewriting the joint density $p(Y_{1,T}, Y_{2,S})$ as the product of the marginal and conditional densities, $p(Y_{1,T}) \cdot p(Y_{2,S}|Y_{1,T})$.[9] In order to state the resulting "combined-sample" estimators, first define the coefficient matrix from a multivariate regression of $R_{2,t}$ on $R_{1,t}$, estimated using the truncated sample,

$$\hat{C} = \begin{bmatrix} \hat{\alpha}' \\ \hat{B}' \end{bmatrix} = (X'X)^{-1} X' Y_{2,S} \ , \tag{11}$$

where $\hat{\alpha}$ is $N_2 \times 1$, $\hat{B}$ is $N_2 \times N_1$, and

$$X = \begin{bmatrix} \iota_S & Y_{1,S} \end{bmatrix}. \tag{12}$$

[9]The details are presented in the Appendix.

The sample residual-covariance matrix from the regression is

$$\hat{\Sigma} = \frac{1}{S}(Y_{2,S} - X\hat{C})'(Y_{2,S} - X\hat{C}). \tag{13}$$

Proposition 1. *Given the likelihood function in (7), the maximum-likelihood estimators of E and V are given by*

$$\hat{E} = \begin{bmatrix} \hat{E}_1 \\ \hat{E}_2 \end{bmatrix}, \tag{14}$$

and

$$\hat{V} = \begin{bmatrix} \hat{V}_{11} & \hat{V}_{21}' \\ \hat{V}_{21} & \hat{V}_{22} \end{bmatrix}, \tag{15}$$

where

$$\hat{E}_1 = \frac{1}{T}\sum_{t=1}^{T} R_{1,t}, \tag{16}$$

$$\hat{E}_2 = \hat{E}_{2,S} + \hat{B}(\hat{E}_1 - \hat{E}_{1,S}), \tag{17}$$

$$\hat{V}_{11} = \frac{1}{T}\sum_{t=1}^{T}(R_{1,t} - \hat{E}_1)(R_{1,t} - \hat{E}_1)', \tag{18}$$

$$\hat{V}_{21} = \hat{B}\hat{V}_{11} \quad (= \hat{V}_{12}') \tag{19}$$

and

$$\hat{V}_{22} = \hat{\Sigma} + \hat{B}\hat{V}_{11}\hat{B}'. \tag{20}$$

Proof: see Anderson (1957) and the Appendix.

It is easily verified that, if $\hat{V}_{11}$ and $\hat{\Sigma}$ are positive definite, then the above combined-sample MLE of the covariance matrix, $\hat{V}$, is positive definite as well.[10] This property is obviously desirable in an estimator of a covariance matrix, but it is not necessarily satisfied by alternative estimators that use the combined sample. For example,

$$V^{\dagger} = \begin{bmatrix} \hat{V}_{11} & \hat{V}_{12,S} \\ \hat{V}_{21,S} & \hat{V}_{22,S} \end{bmatrix} \tag{21}$$

[10]This follows by using (18) through (20) to rewrite $\hat{V}$ as

$$\hat{V} = \begin{bmatrix} I_{N_1} \\ \hat{B} \end{bmatrix} \hat{V}_{11} \begin{bmatrix} I_{N_1} & \hat{B}' \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \hat{\Sigma} \end{bmatrix}.$$

Observe that, for any non-zero real $1 \times N$ vector $z = [z_1\ z_2]$, where the partitioning conforms to that of $\hat{V}$, $z\hat{V}z' \geq z_2\hat{\Sigma}z_2'$ and the latter quadratic form is greater than zero unless z_2 is the zero vector. In that case, $z\hat{V}z' = z_1\hat{V}_{11}z_1' > 0$.

need not be positive definite. Estimators of correlation matrices that use the combined sample can confront similar difficulties. For example, if $\hat{\Psi}_{11}$ denotes the sample correlation matrix constructed from $\hat{V}_{11}$, and $\hat{\Psi}_S$ denotes the (equivalently partitioned) sample correlation matrix constructed from $\hat{V}_S$, then

$$\Psi^{\dagger} = \begin{bmatrix} \hat{\Psi}_{11} & \hat{\Psi}_{12,S} \\ \hat{\Psi}_{21,S} & \hat{\Psi}_{22,S} \end{bmatrix} \tag{22}$$

need not be positive definite. Of course, since $\hat{V}$ is positive definite, the corresponding combined-sample MLE of the correlation matrix is positive definite as well.

The combined-sample MLE's of expected returns can also be interpreted in terms of the Generalized Method of Moments (GMM) of Hansen (1982). Specifically, $\hat{E}$ is also the solution to

$$\min_{E} g'Wg, \tag{23}$$

where

$$g = \begin{bmatrix} \frac{1}{T-S}\sum_{t=1}^{s-1}(R_{1,t} - E_1) \\ \frac{1}{S}\sum_{t=s}^{T}(R_t - E) \end{bmatrix}, \tag{24}$$

and

$$W = \begin{bmatrix} \frac{1}{T-S}\hat{V}_{11,S} & 0 \\ 0 & \frac{1}{S}\hat{V}_S \end{bmatrix}^{-1}. \tag{25}$$

In other words, $\hat{E}$ is the GMM estimator based on the moment conditions in (24) and the weighting matrix in (25). Each diagonal block of W^{-1} corresponds to the usual covariance matrix for a vector of sample means, where the covariances are estimated using the truncated sample, and the zero off-diagonal blocks reflect the assumed temporal independence of returns.[11] The second subvector of g contains N just-identifying conditions that, by themselves, would simply give $\hat{E}_S$ as the GMM estimator. The first subvector of g contains an additional set of N_1 over-identifying conditions for E_1, based on the first $s-1$ observations of $R_{1,t}$, and those over-identifying conditions affect the estimation of both E_1 and E_2.

2.4. The Role of the Longer Histories

As noted earlier, the truncated-sample estimators in (9) and (10) ignore the additional information in the other $T-S$ observations of $R_{1,t}$. Not surprisingly, this additional information is useful in estimating E_1 and V_{11}. More interesting is that this additional information is

[11] It can be shown that $\hat{E}$ is also the GMM estimator when the weighting matrix in (25) is constructed using the combined-sample MLE's $\hat{V}_{11}$ and $\hat{V}$ in place of the truncated-sample estimators $\hat{V}_{11,S}$ and $\hat{V}_S$.

also useful in estimating E_2, V_{22}, and V_{21}. Using the above results, the combined-sample estimators of these quantities can be written as

$$\hat{E}_2 = \hat{E}_{2,S} - \hat{B}(\hat{E}_{1,S} - \hat{E}_1), \tag{26}$$

$$\hat{V}_{22} = \hat{V}_{22,S} - \hat{B}(\hat{V}_{11,S} - \hat{V}_{11})\hat{B}', \tag{27}$$

and

$$\hat{V}_{21} = \hat{V}_{21,S} - \hat{B}(\hat{V}_{11,S} - \hat{V}_{11}). \tag{28}$$

In general, if $R_{1,t}$ and $R_{2,t}$ exhibit nonzero correlations with each other, as reflected in the matrix of estimated regression slopes, $\hat{B}$, then differences between the combined-sample and truncated-sample estimates of the moments of $R_{1,t}$ produce corresponding differences in the estimated moments of $R_{2,t}$.

The basic ideas can be seen most clearly with only two assets ($N_1 = N_2 = 1$), since all of the quantities in equations (26) through (28) are then scalars. The additional information in the first $s-1$ returns on asset 1 enters the estimation of asset 2's expected return in a fairly obvious manner. Suppose, for example, that asset 1 experienced a higher average return during the more recent S periods than over the entire T-period sample, i.e., $\hat{E}_{1,S} - \hat{E}_1 > 0$. The assumed i.i.d behavior for the returns on asset 1, coupled with the information from asset 1's T-period history, implies that the average return over the recent S periods, $\hat{E}_{1,S}$, is too high an estimate of expected return when compared to the value of the more precise estimator, $\hat{E}_1$. If the returns on assets 1 and 2 exhibit positive sample correlation over their common histories, so $\hat{B} > 0$, then $\hat{E}_{2,S}$ is also judged to be too high an estimate of asset 2's expected return, and that truncated-sample estimator is adjusted downward by the amount $\hat{B}(\hat{E}_{1,S} - \hat{E}_1)$. This adjustment follows the same form as the relation,

$$E_2 = \mathrm{E}\{\hat{E}_{2,S}|\hat{E}_{1,S}\} - B(\hat{E}_{1,S} - E_1), \tag{29}$$

implied by the regression function under normality, where $B = V_{21}V_{11}^{-1}$. The right-hand sides of (26) and (29) are similar, with $\mathrm{E}\{\hat{E}_{2,S}|\hat{E}_{1,S}\}$, E_1, and B in (29) replaced by the estimators $\hat{E}_{2,S}$, $\hat{E}_1$, and $\hat{B}$ in (26). Note that such an adjustment could even reverse the relative estimated expected returns on the assets. That is, $\hat{E}_2 - \hat{E}_1$ can have a different sign from $\hat{E}_{2,S} - \hat{E}_{1,S}$.[12]

Asset 1's longer history also provides additional information about the variance of asset 2's return as well as the covariance between returns on the two assets. Suppose, for example, that asset 1 experienced higher volatility during the most recent S periods than over the

[12]If, in the example discussed, $\hat{B} > 1$, then one could observe $\hat{E}_{2,S} - \hat{E}_{1,S} > 0$ but $\hat{E}_2 - \hat{E}_1 < 0$.

entire sample, i.e., $\hat{V}_{11,S} - \hat{V}_{11} > 0$. In other words, $\hat{V}_{11,S}$ is too high an estimate of V_{11} when compared to the value of the more precise estimator, $\hat{V}_{11}$. That information suggests that $\hat{V}_{22,S}$ and $\hat{V}_{21,S}$ are also too high (in absolute value) as estimates of V_{22} and V_{21}. The adjustments in (27) and (28) reflect the property that, if returns on the two assets are correlated, then high ex post variance of $R_{1,t}$ in the most recent S periods is likely to be accompanied by high ex post variance of $R_{2,t}$ and high ex post covariance (in absolute value) between $R_{1,t}$ and $R_{2,t}$. This statement follows from the properties,

$$V_{22} = \text{var}\{R_{2,t} \mid (R_{1,t} - E_1)^2\} - B^2[(R_{1,t} - E_1)^2 - V_{11}], \tag{30}$$

and

$$V_{21} = \text{cov}\{R_{1,t},\ R_{2,t} \mid (R_{1,t} - E_1)^2\} - B[(R_{1,t} - E_1)^2 - V_{11}], \tag{31}$$

which are implied by the joint normality of $R_{1,t}$ and $R_{2,t}$.[13] Note that the relations in (27) and (28) are direct analogs of (30) and (31).

In the two-asset case, if asset 1 experiences higher ex-post variance during the more recent S periods than during its longer history, then the combined-sample maximum-likelihood estimator of the correlation,

$$\hat{\rho}_{12} = \frac{\hat{V}_{21}}{(\hat{V}_{11}\hat{V}_{22})^{1/2}}, \tag{32}$$

is less (in absolute value) than the truncated-sample estimator,

$$\hat{\rho}_{12,S} = \frac{\hat{V}_{21,S}}{(\hat{V}_{11,S}\hat{V}_{22,S})^{1/2}}. \tag{33}$$

Specifically, (27) and (28) imply

$$\hat{\rho}_{12}^2 = \hat{\rho}_{12,S}^2 \left[\hat{\rho}_{12,S}^2 + \frac{\hat{V}_{11,S}}{\hat{V}_{11}}(1 - \hat{\rho}_{12,S}^2)\right]^{-1}, \tag{34}$$

[13] Let $\delta = R_{1,t} - E_1$, and observe that, under normality, $\text{E}\{\delta|\delta^2\} = 0$. Then, by standard rules of variance decomposition,

$$\begin{aligned}
\text{var}\{R_{2,t}|\delta^2\} &= \text{E}\{\text{var}\{R_{2,t}|\delta\}|\delta^2\} + \text{var}\{\text{E}\{R_{2,t}|\delta\}|\delta^2\} \\
&= \text{E}\{V_{22} - B^2V_{11}|\delta^2\} + \text{var}\{E_2 + B\delta|\delta^2\} \\
&= [V_{22} - B^2V_{11}] + B^2\delta^2 \\
&= V_{22} + B^2(\delta^2 - V_{11}),
\end{aligned}$$

and

$$\begin{aligned}
\text{cov}\{R_{1,t},\ R_{2,t}|\delta^2\} &= \text{E}\{\text{cov}\{R_{1,t},\ R_{2,t}|\delta\}|\delta^2\} + \text{cov}\{\text{E}\{R_{1,t}|\delta\},\ \text{E}\{R_{2,t}|\delta\}|\delta^2\} \\
&= \text{E}\{0|\delta^2\} + \text{cov}\{E_1 + \delta,\ E_2 + B\delta|\delta^2\} \\
&= (BV_{11} - BV_{11}) + B\delta^2 \\
&= V_{21} + B(\delta^2 - V_{11}).
\end{aligned}$$

so $\hat{V}_{11,S} - \hat{V}_{11} > 0 \iff \hat{\rho}^2_{12,S} > \hat{\rho}^2_{12}$ (unless $\hat{\rho}_{12,S} = 0$). In other words, the above observations about variance and covariance also apply to the correlation.

3. Portfolio Analysis with Estimation Risk

3.1. The Bayesian Approach

The sample information observed through time T consists of $\Phi_T = \{Y_{1,T}, Y_{2,S}, s\}$, the returns data and the starting period of the short-history assets. Consider an investor with a one-period investment horizon who, after observing this sample, must make an investment decision at the end of period T. It is assumed that the investor finds the historical evidence useful and assesses the characteristics of potential investments in terms of the conditional distribution $p(R_{T+1}|\Phi_T)$. In the multivariate normal setting, if the historical sample were infinitely long, or if the investor somehow otherwise knew the true values of E and V, then $p(R_{T+1}|\Phi_T)$ would simply be the multivariate normal density with those parameters. In practice, the sample Φ_T contains information that is useful to the investor, but, even after observing that sample, the investor does not know the true values of E and V. Thus, part of the risk that the investor rationally perceives arises from parameter uncertainty, or "estimation risk," which would be neglected if the investor were simply to view, say, the maximum-likelihood estimates as if they were the true parameters. Moreover, in the presence of estimation risk, $p(R_{T+1}|\Phi_T)$ is generally not a multivariate normal density.

As illustrated by Zellner and Chetty (1965), Klein and Bawa (1976), and others, portfolio opportunities can be assessed in a Bayesian framework, wherein the conditional distribution $p(R_{T+1}|\Phi_T)$ is obtained using standard Bayesian principles. First consider the case in which s is non-stochastic. Before observing the sample Φ_T, the investor has beliefs about E and V represented by the prior density $p(E, V)$. The prior density is specified here as

$$p(E, V) \propto |V|^{-\frac{N+1}{2}}, \tag{35}$$

which is the standard diffuse prior used to represent "noninformative" beliefs about the parameters of a multivariate normal distribution.[14] The likelihood function in (7) is the density $p(Y_{1,T}, Y_{2,S}|s, E, V)$, and the investor uses this likelihood function along with Φ_T to form updated beliefs about E and V, represented by the posterior density,

$$p(E, V|\Phi_T) \propto p(E, V)\, p(Y_{1,T}, Y_{2,S}|s, E, V). \tag{36}$$

[14] See, for example, Box and Tiao (1973).

When s is stochastic, it is assumed that ζ, the vector of additional parameters in the joint distribution of returns and s, is independent of E and V in the joint prior:

$$p(E, V, \zeta) = p(E, V)\, p(\zeta). \tag{37}$$

In that case, the product of the prior in (37) and the likelihood in (8) gives the joint posterior for E, V, and ζ. Since ζ appears only in the first factor in (8), integrating that joint posterior with respect to ζ gives

$$\begin{aligned} p(E, V|\Phi_T) &= \int_\zeta p(E, V, \zeta|\Phi_T) d\zeta \\ &\propto \int_\zeta p(E, V, \zeta)\, p(Y_{1,T}, Y_{2,S}, s|E, V, \zeta) d\zeta \\ &= \left(\int_\zeta p(\zeta)\, p(s|Y_{1,T}, \zeta) d\zeta\right) p(E, V)\, p(Y_{1,T}, Y_{2,S}|s, E, V) \\ &\propto p(E, V)\, p(Y_{1,T}, Y_{2,S}|s, E, V), \end{aligned} \tag{38}$$

which is the same posterior for E and V as in (36).

To obtain the conditional density $p(R_{T+1}|\Phi_T)$, known as the Bayesian "predictive pdf," the posterior in (36) is first multiplied by $p(R_{T+1}|E, V, \Phi_T)$ to obtain

$$p(R_{T+1}, E, V|\Phi_T) = p(R_{T+1}|E, V, \Phi_T)\; p(E, V|\Phi_T). \tag{39}$$

Integration of the joint density in (39) with respect to E and V then gives the predictive pdf,

$$p(R_{T+1}|\Phi_T) = \int_E \int_V p(R_{T+1}, E, V|\Phi_T) dE\, dV. \tag{40}$$

This predictive pdf can be used to determine the portfolio that satisfies a given investment objective, such as maximizing the expected value of a utility function. The Appendix provides the predictive pdf for R_{T+1} that follows from the prior in (35) and the likelihood function in (7).

3.2. The Mean-Variance Setting

The examples presented in this study are confined to investment objectives involving only the first and second moments of returns. As is well known, a mean-variance characterization of investment opportunities is often a somewhat arbitrary simplification. For example, a mean-variance objective function is not necessarily consistent with expected-utility maximization.[15]

[15]Indeed, except for the case of quadratic utility, a mean-variance objective is likely to provide only an approximation in this framework. In particular, the predictive distribution $p(R_{T+1}|\Phi_T)$, given in the Appendix, does not appear to belong to the class of elliptical distributions, for which mean-variance analysis can be given an expected-utility justification (see Ingersoll (1987)).

A mean-variance framework is used here simply as a familiar setting in which to illustrate the essential aspects of investment analysis when assets' histories differ in length.

These first and second moments of the predictive pdf for R_{T+1},

$$\tilde{E} \equiv \mathrm{E}\{R_{T+1}|\Phi_T\} \tag{41}$$

and

$$\tilde{V} \equiv \mathrm{cov}\{R_{T+1},\ R'_{T+1}|\Phi_T\}, \tag{42}$$

are given in the following proposition.[16]

Proposition 2. *Given the prior density in (35), the likelihood function in (7), and the sample* $\Phi_T = \{Y_{1,T}, Y_{2,S}, s\}$, *then*

$$\tilde{E} = \hat{E} \tag{43}$$

and

$$\tilde{V} = \mathrm{cov}\left\{ \begin{bmatrix} R_{1,t} \\ R_{2,t} \end{bmatrix}, \begin{bmatrix} R'_{1,t} & R'_{2,t} \end{bmatrix} | \Phi_T \right\} = \begin{bmatrix} \tilde{V}_{11} & \tilde{V}_{12} \\ \tilde{V}_{21} & \tilde{V}_{22} \end{bmatrix}, \tag{44}$$

where

$$\tilde{V}_{11} = \left(\frac{T+1}{T-N-2}\right)\hat{V}_{11}, \tag{45}$$

$$\tilde{V}_{12} = \left(\frac{T+1}{T-N-2}\right)\hat{V}_{12} \quad (= \tilde{V}'_{21}), \tag{46}$$

$$\tilde{V}_{22} = \kappa \cdot \hat{\Sigma} + \left(\frac{T+1}{T-N-2}\right)\hat{B}\hat{V}_{11}\hat{B}', \tag{47}$$

$$\kappa = \left(\frac{S}{S-N_2-2}\right)\left(1 + \frac{1}{S}\left[1 + \left(\frac{T+1}{T-N-2}\right)\mathrm{tr}\,(\hat{V}_{11,S}^{-1}\hat{V}_{11}) + (\hat{E}_1 - \hat{E}_{1,S})'\hat{V}_{11,S}^{-1}(\hat{E}_1 - \hat{E}_{1,S})\right]\right), \tag{48}$$

and "tr" *denotes the trace operator. Proof: see Appendix.*

Observe from (43) that estimation risk does not affect expected returns, in the sense that the mean of the predictive pdf is simply the maximum-likelihood estimate of E. Uncertainty about the true expected returns does contribute to the estimation risk incorporated in the

[16]Barry and Brown (1985, pp. 409-410) give moments of the predictive pdf in the case where the true covariance matrix V is known and the prior for E is diffuse. They state that $\tilde{E}$ then contains the sample averages of each asset's return, but such a result would appear to hold only in cases where the histories of returns on the first N_1 assets do not overlap with the histories of the N_2 assets (contrary to the authors' notation) or where all elements of V_{12} are assumed to equal zero.

predictive pdf. Estimation risk is reflected in the covariance matrix of the predictive pdf, in the sense that $\tilde{V}$ exceeds the maximum-likelihood estimator $\hat{V}$ by a positive-definite matrix.[17] Estimation risk also affects the shape of the predictive pdf, in that $p(R_{T+1}|\Phi_T)$ is not a multivariate normal density, although this effect does not enter the mean-variance portfolio setting assumed here.

4. Multiple Starting Dates

Although the analyses in the preceding sections allow an arbitrary number of N assets, each asset's history is assumed to begin at one of only two possible dates. This section generalizes those analyses to include a larger number of J starting dates; the empirical examples presented in the next section includes such a case. Readers who are uninterested in the details of the methodology can skip to the next section.

For $j = 1, \ldots, J$, let the vector $R_{j,t}$ contain the returns on N_j assets in period t, and assume that the overall sample Φ_T includes S_j observations of these returns for periods $T - S_j + 1, \ldots, T$. The assets are ordered such that $S_1 > S_2 > \cdots > S_J$, and we assume, as before, that the first observation of $R_{1,t}$ corresponds to period 1, so $S_1 = T$. The total number of assets is given by $N = \sum_{j=1}^{J} N_j$.

Let the vector $R_{[j],t}$ contain the returns on the first $N_{[j]} = N_1 + N_2 + \cdots + N_j$ assets in period t,

$$R'_{[j],t} = \begin{bmatrix} R'_{1,t} & R'_{2,t} & \cdots & R'_{j,t} \end{bmatrix}, \tag{49}$$

for $t > T - S_j$. As in the previous analysis, it is assumed that, for $j = 1, \ldots, J$, each observation $R_{[j],t}$ is drawn independently from a multivariate normal distribution with

$$\mathrm{E}\left\{R_{[j],t}\right\} = E_{[j]} = \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_j \end{bmatrix}, \tag{50}$$

[17] This follows by using (45) through (47) to obtain

$$\tilde{V} - \hat{V} = \left(\frac{N+3}{T-N-2}\right) \begin{bmatrix} I_{N_1} \\ \hat{B} \end{bmatrix} \hat{V}_{11} \begin{bmatrix} I_{N_1} & \hat{B}' \end{bmatrix} + (\kappa - 1) \begin{bmatrix} 0 & 0 \\ 0 & \hat{\Sigma} \end{bmatrix}.$$

Since $\kappa > 1$, $\tilde{V} - \hat{V}$ is positive definite if $\hat{V}$ is positive definite.

and

$$\operatorname{cov}\left\{R_{[j],t}, R'_{[j],t}\right\} = V_{[j]} = \begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1j} \\ V_{21} & V_{22} & \cdots & V_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ V_{j1} & V_{j2} & \cdots & V_{jj} \end{bmatrix}, \tag{51}$$

where the moments for the entire set of N assets are denoted by $E = E_{[J]}$ and $V = V_{[J]}$.

For the most recent S observations, $S \leq S_j$, define

$$Y_{j,S} = \begin{bmatrix} R'_{j,T-S+1} \\ R'_{j,T-S+2} \\ \vdots \\ R'_{j,T} \end{bmatrix}, \quad Y_{[j],S} = \begin{bmatrix} R'_{[j],T-S+1} \\ R'_{[j],T-S+2} \\ \vdots \\ R'_{[j],T} \end{bmatrix}, \quad j = 1, \ldots, J. \tag{52}$$

Let $s = [s_2, s_3, \ldots, s_J]$ denote the vector of starting periods, where $s_j = T - S_j + 1$. As before, s is permitted to be stochastic, where the joint distribution of returns and s depends on a vector of additional parameters ζ. It is assumed that, conditional on the returns on longer-history assets, the distribution of a starting date for shorter-history assets does not depend on E or V:

$$p(s_{j+1}|Y_{1,T}, \ldots, Y_{j,S_j}, E, V, \zeta) = p(s_{j+1}|Y_{1,T}, \ldots, Y_{j,S_j}, \zeta), \quad j = 1, \ldots, J-1, \tag{53}$$

The joint prior for E, V, and ζ is again assumed to obey the independence property in (37). Under these assumptions, the results given below continue to apply when s is stochastic, for reasons that are straightforward extensions of those given for the two-date analysis in the previous sections.

The likelihood function can be written as

$$\begin{aligned} &p(Y_{j,S_j}, j = 1, \ldots, J|s, E, V) \\ &= \prod_{j=1}^{J} \left[\prod_{t=T-S_j+1}^{T-S_{j+1}} \left(\frac{1}{(2\pi)^{N_{[j]}/2}} |V_{[j]}|^{-1/2} \exp\left\{ -\frac{1}{2} (R_{[j],t} - E_{[j]})' V_{[j]}^{-1} (R_{[j],t} - E_{[j]}) \right\} \right) \right], \end{aligned} \tag{54}$$

where, for notational convenience, $S_{J+1} \equiv 0$.

As before, the analysis is facilitated by a set of regression statistics. A regression of $R_{j,t}$ on $R_{[j-1],t}$, estimated using the most recent S_j observations, produces a coefficient matrix

$$\hat{C}_j = (X'_{[j],S_j} X_{[j],S_j})^{-1} X'_{[j],S_j} Y_{j,S_j} = \begin{bmatrix} \hat{\alpha}'_j \\ \hat{B}'_j \end{bmatrix}, \tag{55}$$

where $\hat{\alpha}_j$ is $N_j \times 1$, $\hat{B}_j$ is $N_j \times (N_1 + N_2 + \cdots + N_{j-1})$, and

$$X_{[j],S_j} = \begin{bmatrix} \iota_{S_j} & Y_{[j-1],S_j} \end{bmatrix}. \tag{56}$$

The disturbance covariance matrix estimated using the fitted residuals is

$$\hat{\Sigma}_j = \frac{1}{S_j}(Y_{j,S_j} - X_{[j],S_j}\hat{C}_j)'(Y_{j,S_j} - X_{[j],S_j}\hat{C}_j). \tag{57}$$

The matrices $\hat{\Sigma}_j$ and $(X'_{[j],S_j}X_{[j],S_j})$ are assumed to be nonsingular, which requires that $S_j > \max(N_j, N_{[j-1]})$. For the Bayesian analysis, it is assumed also that $S_j > N - N_{[j-1]} + 2$, so the requirements for S_j can be summarized as

$$S_j > \max\left(N_{[j-1]},\ N - N_{[j-1]} + 2\right), \quad j = 1, \ldots, J. \tag{58}$$

The statistics $\hat{\alpha}_j$, $\hat{B}_j$, and $\hat{\Sigma}_j$, computed for $j = 2, \ldots, J$, are useful in computing the maximum-likelihood estimates of E and V as well as the moments of the Bayesian predictive pdf of R_{T+1}.

Proposition 3. *Given the likelihood function in (54), the maximum-likelihood estimators of E and V are*

$$\hat{E} = \begin{bmatrix} \hat{E}_1 \\ \hat{E}_2 \\ \vdots \\ \hat{E}_J \end{bmatrix} \quad \text{and} \quad \hat{V} = \begin{bmatrix} \hat{V}_{11} & \hat{V}_{12} & \cdots & \hat{V}_{1J} \\ \hat{V}_{21} & \hat{V}_{22} & \cdots & \hat{V}_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{V}_{J1} & \hat{V}_{J2} & \cdots & \hat{V}_{JJ} \end{bmatrix}, \tag{59}$$

where

$$\hat{E}_1 = \hat{E}_{[1]} = \frac{1}{T}Y'_{1,T}\iota_T, \tag{60}$$

$$\hat{V}_{11} = \hat{V}_{[1]} = \frac{1}{T}(Y_{1,T} - \iota_T\hat{E}'_1)'(Y_{1,T} - \iota_T\hat{E}'_1), \tag{61}$$

and, for $j = 2, \ldots, J$,

$$\hat{E}_j = \hat{\alpha}_j + \hat{B}_j\hat{E}_{[j-1]}, \tag{62}$$

$$\hat{E}_{[j]} = \begin{bmatrix} \hat{E}_{[j-1]} \\ \hat{E}_j \end{bmatrix}, \tag{63}$$

and

$$\hat{V}_{[j]} = \begin{bmatrix} \hat{V}_{11} & \hat{V}_{12} & \cdots & \hat{V}_{1j} \\ \hat{V}_{21} & \hat{V}_{22} & \cdots & \hat{V}_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{V}_{j1} & \hat{V}_{j2} & \cdots & \hat{V}_{jj} \end{bmatrix} = \begin{bmatrix} \hat{V}_{[j-1]} & \hat{V}_{[j-1]}\hat{B}'_j \\ \hat{B}_j\hat{V}_{[j-1]} & \hat{\Sigma}_j + \hat{B}_j\hat{V}_{[j-1]}\hat{B}'_j \end{bmatrix}. \tag{64}$$

Proof: see Appendix.

In addition to the statistics defined above, a set of truncated-sample moments is used in the computation of the moments of the Bayesian predictive pdf. For $j = 2, \ldots, J$, define

$$\hat{E}_{[j-1],S_j} = \frac{1}{S_j} Y'_{[j-1],S_j} \iota_{S_j} \tag{65}$$

and

$$\hat{V}_{[j-1],S_j} = \frac{1}{S_j} (Y_{[j-1],S_j} - \iota_{S_j} \hat{E}'_{[j-1],S_j})' (Y_{[j-1],S_j} - \iota_{S_j} \hat{E}'_{[j-1],S_j}) \,. \tag{66}$$

Proposition 4. *Given the prior density in (35), the likelihood function in (54), and the sample* $\Phi_T = \{Y_{j,S_j}, j = 1, \ldots, J, \; s\}$, *then*

$$\tilde{E} = \mathrm{E}\{R_{T+1} | \Phi_T\} = \hat{E} \tag{67}$$

and

$$\tilde{V} = \mathrm{cov}\{R_{T+1}, R'_{T+1} | \Phi_T\} = \begin{bmatrix} \tilde{V}_{11} & \tilde{V}_{12} & \cdots & \tilde{V}_{1J} \\ \tilde{V}_{21} & \tilde{V}_{22} & \cdots & \tilde{V}_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{V}_{J1} & \tilde{V}_{J2} & \cdots & \tilde{V}_{JJ} \end{bmatrix}, \tag{68}$$

where

$$\tilde{V}_{11} = \tilde{V}_{[1]} = \frac{T+1}{T-N-2} \hat{V}_{11}, \tag{69}$$

and, for $j = 2, \ldots, J$,

$$\begin{aligned} \tilde{V}_{[j]} \equiv \mathrm{cov}\{R_{[j],T+1}, R'_{[j],T+1} | \Phi_T\} &= \begin{bmatrix} \tilde{V}_{11} & \tilde{V}_{12} & \cdots & \tilde{V}_{1j} \\ \tilde{V}_{21} & \tilde{V}_{22} & \cdots & \tilde{V}_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{V}_{j1} & \tilde{V}_{j2} & \cdots & \tilde{V}_{jj} \end{bmatrix} \\ &= \begin{bmatrix} \tilde{V}_{[j-1]} & \tilde{V}_{[j-1]} \hat{B}'_j \\ \hat{B}_j \tilde{V}_{[j-1]} & \kappa_j \hat{\Sigma}_j + \hat{B}_j \tilde{V}_{[j-1]} \hat{B}'_j \end{bmatrix}, \end{aligned} \tag{70}$$

where

$$\begin{aligned} \kappa_j &= \left(\frac{S_j}{S_j - N + N_{[j-1]} - 2} \right) \left(1 + \frac{1}{S_j} \left[1 + \mathrm{tr} \left(\hat{V}^{-1}_{[j-1],S_j} \tilde{V}_{[j-1]} \right) \right. \right. \\ &\quad \left. \left. + (\hat{E}_{[j-1]} - \hat{E}_{[j-1],S_j})' \hat{V}^{-1}_{[j-1],S_j} (\hat{E}_{[j-1]} - \hat{E}_{[j-1],S_j}) \right] \right). \end{aligned} \tag{71}$$

Proof: see Appendix.

5. Example 1: Mean-Variance Efficiency

5.1. The Optimization Problem

Assume that R_t denotes the vector of the returns on the N risky assets in excess of the return on a riskless asset (denoted hereafter as "excess" returns). Let ω denote the N-vector of weights invested in the risky assets, so that the excess return on the investor's overall portfolio p in period $T+1$ is given by

$$R_{p,T+1} = \omega' R_{T+1}. \tag{72}$$

The fraction $1 - \iota_N'\omega$ of the overall portfolio is invested in the riskless asset. The investor's optimal portfolio ω^* is assumed to be the solution to

$$\max_{\omega} \left(\mathrm{E}\{R_{p,T+1}|\Phi_T\} - \frac{A}{2} \mathrm{var}\{R_{p,T+1}|\Phi_T\} \right). \tag{73}$$

Thus, the optimal portfolio is assumed to be mean-variance efficient. Grauer and Hakansson (1993) present evidence suggesting that (73) can provide a reasonable approximation to an expected-utility maximization over short investment horizons.[18] The parameter A will be referred to as the investor's relative risk aversion, defined with respect to the investor's utility of wealth at the end of period $T+1$.[19]

The solution to (73) is easily verified to be

$$\omega^* = \frac{1}{A} \lambda \cdot \gamma, \tag{74}$$

where

$$\gamma = \frac{1}{\iota_N' \tilde{V}^{-1} \tilde{E}} \tilde{V}^{-1} \tilde{E}, \tag{75}$$

and

$$\lambda = \frac{\gamma' \tilde{E}}{\gamma' \tilde{V} \gamma}. \tag{76}$$

As is well known, the N-vector γ contains the weights in the portfolio of risky assets having the maximum Sharpe ratio, the "tangent" portfolio, and λ is the ratio of the tangent

[18] Although such scenarios are not encountered here, Klein and Bawa (1977) observe that, for assets with short enough histories, so that the estimation risk for those assets essentially becomes very large, an expected-utility maximizer will, under certain additional conditions, choose to invest nothing in those assets.

[19] Of course, such a characterization is also only an approximation, given a mean-variance approximation to the expected-utility objective. This point is also discussed by Grauer and Hakansson (1993).

portfolio's expected excess return to its variance, or the "price of risk."[20] The overall optimal portfolio is constructed by investing the fraction λ/A in the tangent portfolio and the fraction $(1 - \lambda/A)$ in the riskless asset.

If all N assets have return histories of the same length T, then

$$\tilde{V} = \left(\frac{T+1}{T-N-2}\right)\hat{V}, \tag{77}$$

which be can obtained using (43)–(47) and then setting $S = T$. In that case, κ in (48) simplifies to[21]

$$\kappa = \frac{T+1}{T-N-2}. \tag{78}$$

With (77), γ in (75) can be rewritten in terms of the maximum-likelihood estimators $\hat{E}$ and $\hat{V}$,

$$\gamma = \frac{1}{\iota_N'\tilde{V}^{-1}\tilde{E}}\tilde{V}^{-1}\tilde{E} = \frac{1}{\iota_N'\hat{V}^{-1}\hat{E}}\hat{V}^{-1}\hat{E}. \tag{79}$$

In other words, when all assets have equal-length histories, allowing for estimation risk does not affect the weights in the tangent portfolio: treating the maximum-likelihood estimates as the true parameters gives the same weights as using the Bayesian predictive distribution. This special case corresponds to the setting in Klein and Bawa (1976), who make the same observation about the irrelevance of estimation risk in computing γ. As those authors explain, allowing for estimation risk simply lowers the fraction invested in the tangent portfolio, since the price of risk in (76) can then be rewritten as

$$\lambda = \frac{\gamma'\tilde{E}}{\gamma'\tilde{V}\gamma} = \left(\frac{T-N-2}{T+1}\right)\frac{\gamma'\hat{E}}{\gamma'\hat{V}\gamma}. \tag{80}$$

In the more general setting, where assets have histories of different lengths, both γ and λ are affected by estimation risk, since $\tilde{V}$ is then no longer simply $\hat{V}$ multiplied by a scalar, as in (77).

5.2. The Sample

The above optimization problem is illustrated here for an asset universe consisting of U.S. Treasury-bills, assumed riskless, and three risky index portfolios ($N = 3$): Standard & Poor's

[20] See, for example, Ingersoll (1987). A portfolio's *Sharpe ratio* is its expected excess return divided by its standard deviation of return. Technically, γ gives the portfolio with the highest *absolute* Sharpe ratio, and it is also assumed that $\iota_N'\tilde{V}^{-1}\tilde{E} \neq 0$.

[21] When $T = S$, then $\hat{E}_1 = \hat{E}_{1,S}$ and $\mathrm{tr}(\hat{V}_{11,S}^{-1}\hat{V}_{11}) = N - N_2$.

composite index (USA), Morgan Stanley Capital International's index for Europe, Australia, and the Far East (EAFE), and the International Finance Corporation's composite index for emerging markets (EMERGE). The returns on each index portfolio are computed as monthly U.S.-Dollar returns in excess of the return on a one-month U.S. Treasury-bill.[22] The IFC emerging-market returns are available beginning in January 1985, whereas the data for the S&P and EAFE indices are available earlier. The EAFE returns are available beginning in January 1970, and, in order to simplify this example, that month is selected as the first observation for returns on the S&P as well, even though returns on the latter index are obviously available well before that date. Data for all three series are included here through December 1995. Thus, in this example, $N_1 = 2$ and $T = 312$, as determined by the sample period of 1/70–12/95 for the S&P and the EAFE indices, while $N_2 = 1$ and $S = 132$, as determined by the sample period of 1/85–12/95 for the emerging-markets index.

5.3. Parameter Estimates

Table 1 reports the means, standard deviations, and correlations for the Bayesian predictive pdf. Maximum-likelihood estimates of those parameters are also reported. As discussed previously, the means of the predictive pdf are identical to the maximum-likelihood estimates. When the truncated sample is used (panel B), the correlations for the predictive pdf are also identical to the maximum-likelihood estimates, since the variance-covariance matrix of the predictive pdf is then simply a scalar multiple of the maximum-likelihood estimate, as given in (77). In that case, the weights in the tangent portfolio γ, shown in the last column of table 1, are the same under the two sets of parameter values, as noted previously (equation (79). With the combined sample (panel A), the correlations from the predictive pdf differ from the maximum-likelihood estimates, so the weights in the tangent portfolio differ as well. In general, however, we see that the parameters and tangent-portfolio weights from the Bayesian predictive pdf are quite close to the maximum-likelihood estimates. In other words, with only three assets, where the shortest history is 132 months, the effects of estimation risk do not appear to be substantial. In general, as will be illustrated in the example presented in the next section, estimation risk becomes more important as the number of assets increases relative to the lengths of the assets' histories.

This example serves primarily to illustrate the potential effects of including additional information provided by the longer-history assets. That is, the differences in table 1 between the combined-sample and truncated-sample results are more substantial than the differences

[22]The data for this study were obtained from CRSP and Datastream.

due to estimation risk. Observe that, for the both the USA and EAFE indices, the truncated period from 1985–95 (panel B) produces higher estimates of mean excess returns than does the longer 1970–95 period (panel A). Next observe that the same statement holds for the emerging-markets index (EMERGE), keeping in mind that the data for that index do not exist before 1985. Incorporating the additional data prior to 1985 results in lower means of USA and EAFE than obtained with the post-1985 data, and, given the positive association between EMERGE and those two indices, the pre-1985 data produce a similar revision in the mean of EMERGE.

The manner by which the pre-1985 data on USA and EAFE supply information about the expected return for EMERGE follows the earlier discussion (section 2) of the differences between the combined-sample and truncated-sample maximum-likelihood estimates in equation (26). Based on the quantities reported in panel B of table 1, it is easily verified that, in a regression of EMERGE on USA and EAFE, the estimated slope coefficients are

$$\hat{B} = \begin{bmatrix} 0.344 & 0.234 \end{bmatrix}. \tag{81}$$

Given that, during the 1985–95 period, $R_{2,t}$ (EMERGE) exhibits this positive association with $R_{1,t}$ (USA and EAFE), the negative differences between the combined-sample and truncated-sample estimates for the means of USA and EAFE produce a corresponding negative difference between the combined-sample and truncated-sample mean for EMERGE,

$$\hat{E}_2 - \hat{E}_{2,S} = -\hat{B}(\hat{E}_{1,S} - \hat{E}_1) = -\begin{bmatrix} 0.344 & 0.234 \end{bmatrix} \begin{bmatrix} 0.55 \\ 0.71 \end{bmatrix} = -0.24, \tag{82}$$

using equation (26).

5.4. Portfolio Implications

Portfolio optimization provides an economic basis for comparing the various methods of estimating the first and second moments of the return distribution. Figure 1 displays the minimum-standard-deviation boundaries for portfolios that combine USA, EAFE, and EMERGE. The higher means for the truncated-sample are evident in the relative positions of the boundaries for that period. For both the truncated and combined samples, the boundaries based on the maximum-likelihood estimates (dashed curve) are close to those based on the Bayesian predictive pdf (solid curve), which again reflects the relatively minor role played by estimation risk in this three-asset example.

The last column of table 1 reports the weights in the tangent portfolio γ implied by the various sets of parameter estimates. In the combined sample, the values for γ based on the

Bayesian predictive pdf differ slightly from those based on maximum-likelihood estimates, but both approaches give portfolio weights of about 30% for USA, 41% for EAFE, and 29% for EMERGE. As noted earlier, when all return series are of the same length, as in the truncated sample, then computing γ using the Bayesian predictive pdf produces the same result as using the ML estimates. In the truncated sample, the weights are 53% for USA, 30% for EAFE, and 17% for EMERGE. Thus, an investor who uses the truncated sample instead of the combined sample would place more weight in USA and less weight in EAFE and EMERGE.

The tangent portfolio possesses the maximum Sharpe ratio within the universe of investments considered. Panel A of table 2 reports the value of the maximum Sharpe measure as computed under the various sets of parameter estimates. The maximum Sharpe ratio is 0.240 (Bayesian) or 0.245 (maximum likelihood) using the truncated-sample estimates, but the maximum Sharpe ratio is only 0.146 (Bayesian) or 0.148 (maximum likelihood) when using the combined-sample estimates. Thus, an investor using the truncated sample would perceive a higher maximum Sharpe ratio than an investor who uses the combined sample. Suppose, however, that we compute Sharpe ratios for all portfolios from the perspective of the latter investor, and that investor also accounts for estimation risk (i.e., uses the Bayesian predictive pdf). Panel B reports the Sharpe ratios perceived by that investor for the tangent portfolios constructed by investors using other samples or estimation methods. Note that the portfolio thought to have a Sharpe ratio of 0.245 by an investor using the truncated-sample MLE's is instead thought to have a Sharpe ratio of only 0.141 by the combined-sample Bayesian investor.

With the mean-variance objective function in (73), the optimal portfolio combines investments in the tangent portfolio and the riskless asset, where, as discussed previously, the proportion in the tangent portfolio is equal to λ/A. (Recall the discussion surrounding equations (74) to (76).) Panel C of table 2 reports this optimal proportion in the tangent portfolio, where the tangent portfolio's composition (γ) and price of risk (λ) are computed using the various samples and estimation methods. Results are presented for three values of A—one, three, and five. Recall from figure 1 that, in this example, the truncated-sample means are substantially higher than the combined-sample means, whereas the truncated-sample volatilities are fairly similar to the combined-sample volatilities. As a result, the optimal proportion in the tangent portfolio is substantially higher when the truncated sample is used. For example, an investor with $A = 5$ invests 128% in the tangent portfolio when using the truncated sample MLE's, whereas an investor with the same risk aversion who instead uses the combined-sample Bayesian predictive pdf invests only 72% in the tangent

portfolio.

The value to the investor of including the pre-1985 data and accounting for estimation risk can be assessed at time T in terms of the objective function in (73). Define the "certainty equivalent" associated with any given portfolio q as

$$C_q = \mathrm{E}\{R_{q,T+1}|\Phi_T\} - \frac{A}{2}\,\mathrm{var}\{R_{q,T+1}|\Phi_T\}. \tag{83}$$

That is, portfolio q achieves the same value for the objective function as does a portfolio providing a riskless excess return of C_q. If the combined-sample Bayesian investor optimally chooses portfolio p, then that investor assigns a certainty-equivalent loss of $C_p - C_q$ to a suboptimal portfolio q. These certainty-equivalent losses, as perceived by the combined-sample Bayesian investor, are reported in panel D of table 2, where the suboptimal portfolio q is constructed using the truncated sample and/or maximum-likelihood estimates.

When the combined sample is used, the certainty-equivalent losses associated with using the maximum-likelihood estimates instead of the Bayesian predictive pdf are very small, less than 0.1 basis points per month. These results are consistent with the earlier observations about the modest role of estimation risk in this three-asset example. Substantially larger losses are associated with portfolios constructed using the truncated sample, ranging from 11 basis points per month ($A = 5$, Bayesian predictive pdf) to over 67 basis points per month ($A = 1$, maximum likelihood). Moreover, when the truncated sample is used, a failure to account for estimation risk adds nontrivially to the certainty-equivalent losses: the differences between the Bayesian and MLE losses range from about 2.3 basis points ($A = 5$) to nearly 12 basis points ($A = 1$). In this example of mean-variance optimization, truncating the sample and then ignoring estimation risk results in a portfolio choice that is rather undesirable when evaluated by an investor who uses the combined sample and accounts for estimation risk.

6. Example 2: Variance Minimization

The previous example illustrates how the the longer histories of some assets can provide useful information about expected returns on all assets. In that example, the estimated covariance matrices of returns are fairly similar across the combined and truncated samples, and, in either sample, estimation risk produces relatively small differences between the covariance matrix of the predictive pdf and the maximum-likelihood estimate of V. This section considers an investment problem in which the optimal portfolio depends only on the

covariance matrix of returns. Moreover, in this example, the estimation risk included in the covariance matrix of the predictive pdf plays an important role in the investment decision.

6.1. The Optimization Problem

As in the previous example, R_t denotes the vector of excess returns on the N risky assets and ω denotes the vector of weights on those assets, so $R_{p,T+1} = \omega' R_{T+1}$ again gives the portfolio's excess return. In this example, the investor is assumed to solve

$$\min_{\omega} \text{var}\{R_{p,T+1}|\Phi_T\}, \tag{84}$$

$$\text{s. t. } \iota_N'\omega = 1. \tag{85}$$

The solution to this problem is the minimum-variance portfolio of the N risky assets—the constraint in (85) excludes the riskless asset from the optimal portfolio. Since

$$\text{var}\{R_{p,T+1}|\Phi_T\} = \omega'\tilde{V}\omega, \tag{86}$$

the solution to the optimization in (84) and (85) is easily verified to be

$$\omega^* = \frac{1}{\iota_N'\tilde{V}^{-1}\iota_N}\tilde{V}^{-1}\iota_N. \tag{87}$$

Thus, unlike the previous example, in which estimated expected returns play a key role, the optimal portfolio in (87) involves only the covariance matrix.

When all N assets have return histories of the same length T, then the resulting simplification of $\tilde{V}$ in (77) allows the solution in (87) to be rewritten with $\hat{V}$ replacing $\tilde{V}$. In other words, estimation risk does not affect the weights in the minimum-variance portfolio when all assets have equal-length histories, which corresponds to the same property for the tangent portfolio observed in the previous section. As also observed there for the tangent portfolio, estimation risk does affect the composition of the minimum-variance portfolio when assets have histories of different lengths.

6.2. The Sample

The above variance-minimization problem is illustrated here using a universe of country-specific index portfolios for 22 emerging markets ($N = 22$). The returns on each country's index are constructed by the International Finance Corporation (IFC) to reflect the portion

of the country's equity market that is accessible to foreign investors.[23] In this example, the returns data for all 22 of these "investable" country portfolios extend through 12/1995. All returns are U.S.-Dollar returns in excess of the one-month U.S. Treasury-bill rate. The first sample month for 10 of the country portfolios is 1/1989; the starting months for the remaining 12 countries range from 9/1989 to 11/1993. Thus, these emerging-market return histories range in length from 84 months to 26 months.[24] Table 3 lists, for each of the 22 countries, the first month of data and the number of observations.

6.3. Parameter Estimates

Table 3 reports, for each country's monthly excess return, the standard deviation computed using five different methods (labeled I through V). Methods I and II use the combined-sample, wherein the lengths of return histories differ across assets. Each of the standard deviations from the Bayesian predictive pdf (method I) exceeds the corresponding maximum-likelihood estimate (method II). The differences, which reflect estimation risk, often run several hundred basis points or more. In the case of Peru, for example, incorporating estimation risk produces nearly a two-thirds increase over the maximum-likelihood estimate of standard deviation (23.4% versus 14.2%). Methods (III) and (IV) use only the most recent 26 months of data for each country, so that each country's return history is truncated to be the same length as the return histories of China and Zimbabwe. When the investor's information about this set of 22 investments is confined to this relatively short period, estimation risk becomes the dominant source of volatility perceived by the investor. The Bayesian predictive standard deviations in that case (method III) are 3.7 times the corresponding maximum-likelihood estimates (method IV).[25] Method V computes, separately for each country, the univariate maximum-likelihood estimate of standard deviation using the history available for each country.

Methods II, IV, and V do not incorporate estimation risk, so differences in estimated volatilities across these methods simply reflect differences in ex post variances (and covariances) across the various sample periods. By construction, methods II and V produce identical estimates for the first 10 countries (Argentina through Thailand), which all have return histories of 84 months. For the remaining 12 countries, which have shorter return

[23] See International Finance Corporation (1993).

[24] A few countries in the IFC universe with even shorter histories were excluded because their inclusion would have produced violations of (58).

[25] Recall that, when all assets have histories of length T, the difference between the covariance matrices from the two methods is given by (77).

histories, method II produces higher estimated volatilities than method V in all but one case (Turkey). Similarly, for the 12 shorter-history countries, method II also produces higher estimates than method IV in all but one case (Taiwan). For the shorter-history assets, the higher estimates produced by method II reflect information about volatility provided by the longer-history assets. Many of those longer-history assets experienced less ex post variance during the more recent years than during the earlier years. This general pattern can be seen in a comparison of the estimates from methods II and IV for the longer-history assets. Thus, for many of those assets, the ex post variance of the more recent years is too low an estimate of true variance when compared to the estimate based on the total period. Given that the returns on many of the 22 countries exhibit positive correlations with each other (over periods of common recent history), the ex post variances of the shorter-history assets are also judged to be too low as estimates of the true volatities. This reasoning, which is necessarily fuzzy with many assets and start dates, follows the more precise argument given earlier in section 2 for the two-asset case.

The upper-right portion of table 4 displays the correlations ($\times 100$) based on the combined-sample Bayesian predictive pdf. The lower-left portion displays, for each pair of countries, the difference ($\times 100$) between the Bayesian predictive correlation in the top portion and the bivariate truncated-sample maximum-likelihood estimate computed using the jointly available history for a given pair (so the length of the joint history is equal to the shorter of the two countries.) A simple approach to estimating the variance-covariance matrix might be to combine the latter "available-history" correlation estimates with the variance estimates based on each country's available history (reported under method V in table 3). Aside from other properties of such an approach, one potential problem is that the correlation matrix estimated in this fashion, and thus the corresponding covariance matrix, can fail to be positive definite.[26] Indeed, that is the case in this example. Thus, this approach is not included here among those used to construct the minimum-standard-deviation boundary or the global minimum-variance portfolio.[27] The differences between the combined-sample Bayesian predictive correlations and the available-history correlation estimates are equal to zero for all pairs from the ten countries (Argentina through Thailand) that have data beginning in 1/89, the earliest month of the combined sample. Recall from (77) that, for those countries, the Bayesian predictive covariance matrix is simply a scalar multiple of the maximum-likelihood estimate of the covariance matrix, and thus the correlations, are indentical under both meth-

[26]This "available-history" estimator of the correlation matrix is essentially that given in (22), generalized to multiple starting dates.

[27]If the symmetric matrix A is not positive definite, then the solution to $\min_\omega \omega' A\omega \; s.t. \; \omega'\iota = 1$ need not exist and, in general, is not given by $\omega^* = (1/\iota' A^{-1}\iota)A^{-1}\iota$.

ods. For many of the remaining assets, especially those with the shorter histories, the values in the lower-left portion of table 4 are negative, indicating that the Bayesian predictive correlation is less than the maximum-likelihood estimate based on the jointly available history.

6.4. Portfolio Implications

The rightmost three columns of table 3 display the weights in the minimum-variance portfolio, where the covariance matrix is estimated using methods I through IV. Methods I and II both use the combined sample, but the differences in weights between these two methods reveal that the estimation risk not incorporated in the maximum-likelihood estimates (method II) plays a significant role in computing ω^* in (87). In this example, the weights based on the Bayesian predictive pdf (method I) take less extreme values than the weights based on the maximum-likelihood estimates. The Bayesian weights range from -14% to 45%, with only one weight exceeding 25% in absolute value, whereas the maximum-likelihood weights range from -53% to 82%, and 12 of the weights exceed 25% in absolute value. Estimation risk does not affect the weights computed using the truncated sample (methods III and IV). In that case, the Bayesian predictive covariance matrix is simply a scalar multiple of the maximum-likelihood estimate, as explained previously, and the solution in (87) is unaffected by a scalar multiplication of $\tilde{V}$. The weights produced here by methods III and IV also take more extreme values than those in method I. In this example, those truncated-sample weights happen to resemble fairly closely the weights produced by method II.

Table 5 reports the global minimum standard deviation computed using the various methods. For each method, the standard deviation is computed two ways. The first, shown in panel A, computes the minimum standard deviation using the covariance matrix obtained under the given method. For example, suppose the combined-sample maximum-likelihood estimator $\hat{V}$ is used (in place of $\tilde{V}$) to compute the minimum-variance portfolio (87), and the vector of resulting weights is denoted $\hat{\omega}$ (given earlier in table 3). Then $(\hat{\omega}'\hat{V}\hat{\omega})^{1/2}$ equals 2.31%, as reported in the second column of panel A.

Each value in panel A of table 5 corresponds to the leftmost point on the minimum-standard-deviation boundary constructed with the moments obtained by the given method. These boundaries are displayed in figure 2.[28] When maximum-likelihood estimators are used, the boundary based on the combined sample (II) lies close to that based on the truncated

[28]Note that, although estimated expected returns are not used elsewhere in this example, they are used here to plot the boundaries.

sample (IV), but the leftmost points of both boundaries lie at least 150 basis points to the left of the minimum standard deviation of 3.8% for the combined-sample Bayesian predictive pdf (I). Thus, in this example, estimation risk has a larger effect on volatility than does the inclusion of the additional data in the combined sample. When only the truncated sample is used and estimation risk is incorporated, then the resulting minimum-standard-deviation boundary (III) lies quite far to the right, with a global minimum standard deviation of about 7.5% per month. Unlike the minimum-standard deviation boundaries computed in the previous example (figure 1), the vertical locations of all four boundaries in this example are similar. In other words, whereas the first example served principally to illustrate how perceived portfolio opportunities can be affected by differences across methods in estimating expected returns, such differences exert less influence on the opportunity sets constructed in this example.

Panel B of table 5 displays the standard deviations of the same portfolios constructed for panel A, but the standard deviation of each portfolio is now computed from the perspective of the combined-sample Bayesian investor. If, for example, $\hat{\omega}$ still denotes the vector of weights obtained when $\hat{V}$ is used in (87), then the value in the second column of panel B, 6.09%, is equal to $(\hat{\omega}'\tilde{V}\hat{\omega})^{1/2}$. The results in panel B again reveal the dominant role of estimation risk in this example. To the combined-sample Bayesian investor, the global minimum standard deviation is less than 4 percent, but the portfolios constructed using the other three methods have standard deviations between 6 and 7 percent. This second example differs in many respects from the first but reaches a similar overall conclusion: the portfolios constructed by the other methods are viewed as substantially suboptimal by an investor who uses the combined sample and incorporates estimation risk.

7. Performance in Repeated Samples

The examples in the previous two sections illustrate the conditional Bayesian decision approach, wherein the predictive pdf of returns, and thus the investor's portfolio decision, are conditioned on the single observed sample. As demonstrated in the examples, the predictive pdf can be used by the conditional Bayesian investor to assess the relative merits of various alternative portfolios, such as portfolios formed by methods that truncate the sample or ignore estimation risk.

Another approach to comparing portfolios formed by various methods is to view each method's portfolio selection as a function of the sample and then to compare the perfor-

mances of the methods across repeated random samples. In that approach, the typical performance of each method across repeated samples is computed based on one or more assumed true sets of return moments, a computation that essentially yields the frequentist "risk" function.[29] In practice, the true moments of returns are unknown, and an investor engaged in asset allocation might observe only one sample per lifetime (although that sample would get updated). Nevertheless, studies of portfolio-selection methods often report repeated-sample comparisons, and such an analysis is included here in order to provide a broader perspective on the proposed methodology.[30]

This section reports two repeated-sample experiments, each corresponding to one of the two examples presented earlier. In each experiment, the starting periods (s) for the shorter-history assets are held constant across the randomly generated samples. In the first experiment, excess returns for three risky assets are generated by a multivariate normal distribution whose moments are set equal to the combined-sample maximum-likelihood estimates reported in table 1. Each generated hypothetical sample of monthly returns has the property that, as in the actual sample in example 1, the first two assets have 312 observations (corresponding to the period 1/70–12/95 for USA and EAFE), whereas the third asset has only 132 monthly observations (corresponding to the period 1/70–12/95 for EMERGE). For each generated sample, the weights in the optimal portfolio are computed under each of the four methods analyzed previously (cf. table 2), but the certainty equivalent for each portfolio is computed based on the assumed true E and V used in generating the returns. For example, let ω^* denote the weights in the optimal combined-sample Bayesian portfolio p, which is the solution to (73) where Φ_T denotes the generated sample. Then, rather than using the predictive pdf to compute the certainty equivalent for ω^*, as in (83), the certainty equivalent excess return is instead computed as

$$\begin{aligned} C_p &= \mathrm{E}\{R_{p,t}\} - \frac{A}{2}\mathrm{var}\{R_{p,t}\} \\ &= \omega^{*'}E - \frac{A}{2}\omega^{*'}V\omega^*, \end{aligned} \tag{88}$$

where E and V denote the assumed true moments. In this experiment, relative risk aversion (A) is set equal to 3. These calculations are repeated in each of 5000 independently generated samples.

[29]Let $\omega(\Phi)$ denote a portfolio decision rule, a function of the sample Φ, and let $L(\theta,\omega(\Phi))$ denote the loss associated with a given sample Φ and given parameter vector θ. The portfolio rule's risk function $r(\theta)$, defined on the parameter space Θ, is given by $r(\theta) = \mathrm{E}\{L(\theta,\omega(\Phi))\}$, where the expectation is taken with respect to the distribution of Φ, given θ. Berger (1985) compares approaches based on frequentist risk to those based on conditional Bayesian decision principles.

[30]Previous studies that investigate the frequentist risk of various portfolio-selection methods include Brown (1979), Jorion (1986), and Frost and Savarino (1986).

The results of the first experiment are summarized in table 6. Panel A reports, for each of the four methods, the certainty-equivalent loss relative the optimal portfolio constructed using the true E and V. These results indicate, in a sense, the extent to which an investor loses by not knowing the true E and V. We see that, although such losses are fairly substantial across all methods, the combined-sample Bayesian method typically produces the smallest losses. The mean loss for that approach is about 24 basis points (bp) per month, and mean losses for the other methods range up to about 44 bp, for the truncated-sample maximum-likelihood method. In panel B, the certainty-equivalent loss is computed relative to the combined-sample Bayesian portfolio. That is, the loss is the certainty equivalent for that portfolio minus the certainty equivalent for the portfolio based on one of the other three methods, where the certainty equivalents for both portfolios are computed as in (88), again using the true E and V. The mean loss for the combined-sample maximum-likelihood method is less than 2 bp, but the two truncated-sample methods have mean losses of 15.5 bp (Bayesian) and 19.4 bp (maximum likelihood). Thus, as observed previously for the predictive pdf based on the actual data, ignoring estimation risk in this example is not as costly as truncating the sample. In fact, the mean losses reported in panel B, which are based on certainty equivalents computed with the assumed true moments, are quite similar to the certainty-equivalent losses reported in panel D of table 2, which are based on the Bayesian predictive pdf for the actual sample.

The second experiment corresponds to the variance-minimization in example 2. Returns are generated on 22 assets, and the lengths of histories for the assets are the same as those of the emerging-market country indices in that example. As in the first experiment, 5000 independent samples are generated from a multivariate normal distribution, and the true E and V for the 22 assets are assumed to be equal to the combined-sample maximum-likelihood estimates for the actual data. For each generated sample, the weights in the minimum-variance portfolio are computed using the various methods, and then the standard deviations of the portfolios are computed using the assumed true covariance matrix V. Panel A of table 7 reports the differences between these standard deviations and the true minimum standard deviation. The combined-sample Bayesian portfolio has a standard deviation that, on average, exceeds the true minimum standard deviation by 2.91%. In contrast, the combined-sample maximum-likelihood portfolio's standard deviation exceeds the true minimum by 4.23%, and the corresponding difference for the truncated-sample portfolio is 3.77%. (Recall that, for the truncated sample, the Bayesian and maximum-likelihood weights are identical.) Panel B of table 7 reports the difference between the standard deviation of the portfolio constructed by the indicated method minus the standard deviation of the combined-sample Bayesian

portfolio, where both standard deviations are based on the assumed true covariance matrix V. The mean differences are 1.31% for the combined-sample maximum-likelihood portfolio and 0.86% for the truncated-sample methods. Thus, as observed previously for the predictive pdf based on the actual data, failure to account for estimation risk, even when the combined sample is used, results in a substantially higher volatility.

8. Concluding Remarks

When some assets have shorter return histories than others, it is neither necessary nor desirable to truncate the sample so that the lengths of all return series are determined by the length of the shortest series. In general, the data in a longer-history asset can provide information about the parameters of that asset's returns as well as the parameters of other assets' returns. This point is illustrated here in the context of an i.i.d. multivariate Normal model, but it is likely that the same concept can be demonstrated in other stochastic settings, such as where conditional first or second moments fluctuate through time.

The basic factorization approach exploited here in deriving closed-form analytic results (see Appendix) requires that the time periods covered by the various series can be arranged as nested subsets.[31] When this nesting property fails, such as when one series has both an early starting date and an early ending date, then the return moments can be obtained numerically using data-augmentation methods, such as the E-M algorithm (to obtain maximum-likelihood estimates) or the Gibbs sampler (to obtain the Bayesian predictive pdf).[32] With more complicated stochastic settings, analytical results could be difficult to obtain at all, whether or not the series are nested, and these numerical approaches could then be useful in general.

The concept of using the combined (non-truncated) sample could also be extended to the problem of making inferences about a pricing model or a given portfolio's mean-variance efficiency. In a frequentist setting, the likelihood function employed here could also be used to construct a likelihood-ratio test (LRT) of the efficiency of a given portfolio, where the parameter restrictions are the same as those investigated in previous studies. For example, Gibbons, Ross, and Shanken (1989) derive the finite-sample distribution of the LRT statistic when all assets have equal-length histories; the finite-sample behavior of the LRT in the case

[31] See Little and Rubin (1987) for a deeper discussion of maximum-likelihood estimation in this case, where those authors use the term "monotone data" to denote the nested-subset property.

[32] See Tanner (1993), for example, for a discussion of such methods.

of unequal-length histories presents a topic for future research.

Although the analysis and examples presented here rely on the diffuse prior distribution for E and V in (35), an extension to an informative natural-conjugate prior distribution is straightforward. A natural-conjugate prior for E and V can be interpreted as the posterior distribution obtained by updating the diffuse prior using T_0 hypothetical observations of R_t. Those hypothetical observations produce sample moments $\bar{E}$ and $\bar{V}$, where T_0, $\bar{E}$, and $\bar{V}$ are the parameters to be specified in the prior. The posterior distribution obtained by combining this natural-conjugate prior with the actual data is the same as the posterior distribution obtained by combining the diffuse prior with a sample that appends the hypothetical T_0 observations to the actual data. In other words, the actual data, where the N asset histories can have unequal lengths, is augmented by an additional set of N histories of equal length T_0. The relevant first and second sample moments from this new combined sample can be computed from the moments of the actual and hypothetical data. In the case of two start dates, for example, $\hat{E}_1$ is simply replaced by a weighted average of $\hat{E}_1$ and $\bar{E}_1$, where the respective weights are $T/(T+T_0)$ and $T_0/(T+T_0)$. Similarly, $\hat{E}_S$ is replaced by a weighted average of $\hat{E}_S$ and $\bar{E}$, where the weights are $S/(S+T_0)$ and $T_0/(S+T_0)$.[33]

In a Bayesian setting, the posterior distribution of the parameters of the return distribution (given in the Appendix) could be used to obtain the posterior distribution of a given portfolio's degree of mean-variance inefficiency. Studies by Shanken (1987), Harvey and Zhou (1990), and Kandel, McCulloch, and Stambaugh (1995) investigate this problem in samples where all assets have histories of equal length. When one selects ex ante a portfolio whose degree of inefficiency is of particular interest, then the diffuse prior should probably be replaced by an informative prior constructed with attention given to the implied prior beliefs about the degree of inefficiency in the selected portfolio. Otherwise, as demonstrated by Kandel, McCulloch, and Stambaugh (1995), the implied prior beliefs about any given portfolio are concentrated toward gross inefficiency, such that a very large sample is required in order to infer that any portfolio is close to being efficient.

[33]The expressions for the second moments, as well as extensions to multiple starting dates, are more complicated, but they involve only the statistics from the actual data required in the diffuse-prior analysis and the parameters T_0, $\bar{E}$, and $\bar{V}$.

APPENDIX

This appendix derives the Bayesian predictive pdf of R_{T+1} as well as the first and second moments of that distribution. Proofs are given for the general setting with J starting dates (proposition 4); the result for two starting dates follows directly as a special case (proposition 2). Also included is a summary of Anderson's (1957) method for computing maximum-likelihood estimates (proposition 1), which is straightforward to generalize to the setting with multiple starting dates (proposition 3). The change of variables employed in that method also facilitates the the derivation of the Bayesian results.

For $j = 2, \ldots, J$, define the change of variables

$$B_j = \left[\begin{array}{cccc} V_{j1} & V_{j2} & \cdots & V_{j,j-1} \end{array} \right] V_{[j-1]}^{-1} , \tag{A.1}$$

$$\alpha_j = E_j - B_j E_{[j-1]} , \tag{A.2}$$

and

$$\Sigma_j = V_{jj} - B_j V_{[j-1]} B_j' . \tag{A.3}$$

Let θ denote the vector of original parameters in E and V, and let ξ denote the vector of parameters in E_1, V_{11} and $\{\alpha_j, B_j, \Sigma_j; j = 2, \ldots, J\}$. The vectors θ and ξ have equal numbers of elements, and the Jacobian of the transformation is given by

$$\begin{aligned} \left| \frac{\partial \theta}{\partial \xi'} \right| &= |V_{11}|^{N_2} |V_{[2]}|^{N_3} \cdots |V_{[J-1]}|^{N_J} \\ &= |V_{11}|^{N_J + N_{J-1} + \cdots + N_2} |\Sigma_2|^{N_J + N_{J-1} + \cdots + N_3} \cdots |\Sigma_{J-2}|^{N_J + N_{J-1}} |\Sigma_{J-1}|^{N_J} \\ &= |V_{11}|^{N - N_1} |\Sigma_2|^{N - N_{[2]}} \cdots |\Sigma_{J-2}|^{N - N_{[J-2]}} |\Sigma_{J-1}|^{N - N[J-1]}, \end{aligned} \tag{A.4}$$

which can be verified using equations (A.1) through (A.3) and the relation

$$\begin{aligned} |V_{[j]}| &= |V_{[j-1]}| \, |V_{jj} - B_j V_{[j-1]} B_j'| \\ &= |V_{[j-1]}| \, |\Sigma_j|, \end{aligned} \tag{A.5}$$

for $j = 2, \ldots, J$, where the first equality in (A.5) uses (51) and (A.1) and applies a standard result for the determinant of a partitioned matrix (e.g., Anderson (1984, theorem A.3.2)), and the second equality uses (A.3). The relation in (A.5) can also be used to write

$$|V| = |V_{11}| \, |\Sigma_2| \cdots |\Sigma_J|. \tag{A.6}$$

Using (A.4) and (A.6), the prior for θ in (35) is translated into a prior for ξ given by

$$\begin{aligned} p(\xi) &= p(\theta)\left|\frac{\partial\theta}{\partial\xi'}\right| \\ &\propto |V|^{-\frac{N+1}{2}}|V_{11}|^{N-N_1}|\Sigma_2|^{N-N_{[2]}}\cdots|\Sigma_{J-2}|^{N-N_{[J-2]}}|\Sigma_{J-1}|^{N-N[J-1]} \\ &= |V_{11}|^{\frac{1}{2}(N-2N_1-1)}\prod_{j=2}^{J}|\Sigma_j|^{\frac{1}{2}(N-2N_{[j]}-1)}. \end{aligned} \tag{A.7}$$

Following Anderson (1957), the change of variables in (A.1) through (A.3) allows the likelihood function in (54) to be rewritten as

$$\begin{aligned} &p(Y_{j,S_j}, j=1,\ldots,J|s,\xi) \\ &= p(Y_{1,T}|\xi)\prod_{j=2}^{J}p(Y_{j,S_j}|Y_{[j-1],S_{j-1}}, s_j, \xi) \\ &= p(Y_{1,T}|E_1, V_{11})\prod_{j=2}^{J}p(Y_{j,S_j}|Y_{[j-1],S_j}, s_j, \alpha_j, B_j, \Sigma_j) \\ &= \frac{1}{(2\pi)^{N_1/2}}|V_{11}|^{-\frac{T}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\,(Y_{1,T}-\iota_T E_1')'(Y_{1,T}-\iota_T E_1')V_{11}^{-1}\right\} \times \\ &\quad \prod_{j=2}^{J}\frac{1}{(2\pi)^{N_j/2}}|\Sigma_j|^{-\frac{S_j}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\,(Y_{j,S_j}-X_{[j],S_j}C_j)'(Y_{j,S_j}-X_{[j],S_j}C_j)\Sigma_j^{-1}\right\}, \end{aligned} \tag{A.8}$$

where

$$C_j = \begin{bmatrix} \alpha_j' \\ B_j' \end{bmatrix}. \tag{A.9}$$

Standard results for the multivariate normal model imply that maximizing the first factor in (A.8) with respect to E_1 and V_{11} gives the solutions in (60) and (61). Similarly, standard results for the normal multivariate regression model imply that, for $j=2,\ldots,J$, maximizing the j^{th} factor in (A.8) with respect to C_j and Σ_j gives the solutions in (55) and (57). Reversing the change of variables in (A.1) through (A.3) then gives the maximum-likelihood estimators in (62) through (64).

Both the prior in (A.7) and the likelihood function in (A.8) are expressed as products of J factors, where any given element of ξ appears in only one factor. Therefore, the posterior distribution for ξ exhibits a similar property:

$$\begin{aligned} p(\xi|\Phi_T) &\propto p(\xi)p(Y_{j,S_j}, j=1,\ldots,J|s,\xi) \\ &\propto p(E_1, V_{11}|Y_{1,T})\prod_{j=2}^{J}p(C_j, \Sigma_j|Y_{j,S_j}, Y_{[j-1],S_j}, s_j), \end{aligned} \tag{A.10}$$

where

$$p(E_1, V_{11}|Y_{1,T}) \propto |V_{11}|^{-\frac{L_1+N_1+1}{2}} \exp\left\{-\frac{1}{2}\text{tr}\,(Y_{1,T} - \iota_T E_1')'(Y_{1,T} - \iota_T E_1')V_{11}^{-1}\right\}, \quad \text{(A.11)}$$

$$p(C_j, \Sigma_j|Y_{j,S_j}, Y_{[j-1],S_j}, s_j) \propto |\Sigma_j|^{-\frac{L_j+N_j+1}{2}} \exp\left\{-\frac{1}{2}\text{tr}\,(Y_{j,S_j} - X_{[j],S_j}C_j)'(Y_{j,S_j} - X_{[j],S_j}C_j)\Sigma_j^{-1}\right\}, \quad j = 2, \ldots, J, \quad \text{(A.12)}$$

and

$$L_j = S_j - N + 2N_{[j]} - N_j, \quad j = 1, \ldots, J. \quad \text{(A.13)}$$

A useful property of the factorization in (A.10) is that each of the posterior distributions in (A.11) and (A.12) is easily analyzed in a standard setting. For example, it is straightforward to verify that, for $j = 2, \ldots, J$, (A.12) can be rewritten as

$$p(C_j, \Sigma_j|Y_{j,S_j}, Y_{[j-1],S_j}, s_j) \propto |\Sigma_j|^{-\frac{L_j+N_j+1}{2}} \exp\left\{-\frac{1}{2}\text{tr}\,[Q_j + (C_j - \hat{C}_j)'A_j(C_j - \hat{C}_j)]\Sigma_j^{-1}\right\}, \quad \text{(A.14)}$$

where

$$A_j = X_{[j],S_j}'X_{[j],S_j} \quad \text{(A.15)}$$

and

$$Q_j = S_j\,\hat{\Sigma}_j. \quad \text{(A.16)}$$

The right-hand side of (A.14) is identical to the posterior distribution for C_j and Σ_j in the standard multivariate regression model where a sample of length L_j generates (i) a matrix of cross-products of the independent variables equal to A_j, (ii) a matrix of least-squares coefficient estimates equal to $\hat{C}_j$, and (iii) a matrix of cross-products of fitted residuals equal to Q_j.[35] Therefore, known results for that standard model imply that the predictive pdf for $R_{j,T+1}$, conditional on $R_{[j-1],T+1}$, is a multivariate Student t density:[36]

[35]See Zellner (1971), pp. 224–227. The diffuse prior used in that standard model is

$$p(C, \Sigma_j) \propto |\Sigma_j|^{-\frac{N_j+1}{2}},$$

whereas, from (A.7), the marginal prior on those parameters is

$$p(C, \Sigma_j) \propto |\Sigma_j|^{\frac{N-2N_{[j]}+1}{2}}.$$

[36]See Zellner (1971), pp. 233–236.

$$
\begin{aligned}
& p(R_{j,T+1}|R_{[j-1],T+1}, \Phi_T) \\
& \qquad = \; p(R_{j,T+1}|R_{[j-1],T+1}, Y_{j,S_j}, Y_{[j-1],S_j})) \\
& \qquad \propto \; [\nu_j + (R_{j,T+1} - \hat{\mu}_{j,T+1})' G_j (R_{j,T+1} - \hat{\mu}_{j,T+1})]^{-(N_j+\nu_j)/2}, \qquad \text{(A.17)}
\end{aligned}
$$

where

$$\hat{\mu}_{j,T+1} = \hat{C}_j' x_{[j],T+1}, \qquad \text{(A.18)}$$

$$G_j = [1 - x_{[j],T+1}'(A_j + x_{[j],T+1} x_{[j],T+1}')^{-1} x_{[j],T+1}] \nu_j Q_j^{-1}, \qquad \text{(A.19)}$$

$$x_{[j],T+1}' = [1 \; R_{[j-1],T+1}'], \qquad \text{(A.20)}$$

and

$$
\begin{aligned}
\nu_j &= L_j - N_{[j-1]} - N_j \\
&= S_j - N + N_{[j-1]}. \qquad \text{(A.21)}
\end{aligned}
$$

The first two moments of the above conditional distribution are given by properties of the multivariate t distribution:[37]

$$
\begin{aligned}
\mathrm{E}\{R_{j,T+1}|R_{[j-1],T+1}, \Phi_T\} &= \hat{\mu}_{j,T+1} \\
&= \hat{\alpha}_j + \hat{B}_j R_{[j-1],T+1} \qquad \text{(A.22)}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{cov}\{R_{j,T+1}, \; R_{j,T+1}'|R_{[j-1],T+1}, \Phi_T\} &= \frac{\nu_j}{\nu_j - 2} G_j^{-1} \\
&= \frac{S_j}{\nu_j - 2} \left(1 + \frac{1}{S_j}[1 + u_j]\right) \hat{\Sigma}_j, \qquad \text{(A.23)}
\end{aligned}
$$

where

$$u_j = (R_{[j-1],T+1} - \hat{E}_{[j-1],S_j})' \hat{V}_{[j-1],S_j}^{-1} (R_{[j-1],T+1} - \hat{E}_{[j-1],S_j}). \qquad \text{(A.24)}$$

Similarly, the posterior for E_1 and V_{11} in (A.11) can be rewritten as

$$p(E_1, V_{11}|Y_{1,T}) \;\propto\; |V_{11}|^{-\frac{L_1+N_1+1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\,[Q_1 - (E_1 - \hat{E}_1) A_1 (E_1 - \hat{E}_1)'] V_{11}^{-1}\right\} \qquad \text{(A.25)}$$

where

$$A_1 = T \qquad \text{(A.26)}$$

[37] See Zellner (1971), pp. 331-332 and page 383.

and

$$Q_1 = T\hat{V}_{11}. \tag{A.27}$$

As in the previous case, the posterior in (A.25) is identical to that obtained in the standard multivariate regression model in which a sample of length L_1 generates (i) a matrix of cross-products of the independent variables equal to A_1, (ii) a matrix of least-squares coefficient estimates equal to $\hat{E}_1'$, and (iii) a matrix of cross-products of fitted residuals equal to Q_1.[38] Therefore, following the same analysis as before, we obtain

$$p(R_{1,T+1}|\Phi_T) \propto [\nu_1 + (R_{1,T+1} - \hat{E}_1)'G_1(R_{1,T+1} - \hat{E}_1)]^{-(N_1+\nu_1)/2}, \tag{A.28}$$

where

$$G_1 = [1 - (T+1)^{-1}]\nu_1 Q_1^{-1}, \tag{A.29}$$

$$\nu_1 = L_1 - N_1, \tag{A.30}$$

and the first two moments of this predictive distribution are given by

$$\tilde{E}_1 = \mathrm{E}\{R_{1,T+1}|\Phi_T\} = \hat{E}_1 \tag{A.31}$$

and

$$\begin{aligned} \tilde{V}_{11} = \mathrm{cov}\{R_{1,T+1},\ R_{1,T+1}'|\Phi_T\} &= \frac{\nu_1}{\nu_1 - 2}G_1^{-1} \\ &= \frac{T+1}{T-N-2}\hat{V}_{11}. \end{aligned} \tag{A.32}$$

The predictive pdf for R_{T+1} can be factored as

$$p(R_{T+1}|\Phi_T) = p(R_{1,T+1}|\Phi_T)\prod_{j=2}^{J} p(R_{j,T+1}|R_{[j-1],T+1}, \Phi_T), \tag{A.33}$$

so this density can be obtained simply by multiplying the densities in (A.17) and (A.28).[39] This joint density's first and second moments can be obtained progressively. At each step

[38] In this case, there is only one independent variable, and that variable does not involve $R_{j,t}$ for $j > 1$. The diffuse prior used in the standard model is

$$p(E_1, V_{11}) \propto |V_{11}|^{-\frac{N_1+1}{2}},$$

whereas, from (A.7), the marginal prior on those parameters is

$$p(E_1, V_{11}) \propto |V_{11}|^{\frac{N-2N_1+1}{2}}.$$

[39] The product of the normalizing constants is equal to

$$\left(\frac{\nu_1^{\nu_1/2}\Gamma[(\nu_1+N_1)/2]|G_1|^{1/2}}{\pi^{N_1/2}\Gamma(\nu_1/2)}\right)\prod_{j=2}^{J}\left(\frac{\nu_j^{\nu_j/2}\Gamma[(\nu_j+N_j)/2]|G_j|^{1/2}}{\pi^{N_j/2}\Gamma(\nu_j/2)}\right)$$

j, for $j = 2, \ldots, J$, the moments of $R_{j,T+1}$ in (A.22) and (A.23), which are conditioned on $R_{[j-1],T+1}$ as well as Φ_T, are combined with the moments of $R_{[j-1],T+1}$ that are conditioned only on Φ_T. In the first step, where $j = 2$, the moments of $R_{1,T+1}$ conditioned on Φ_T are given in (A.31) and (A.32). Applying this approach to obtain $\tilde{E}_j$, the mean of $R_{j,T+1}$, gives

$$\begin{aligned} \mathrm{E}\{R_{j,T+1}|\Phi_T\} &= \mathrm{E}\{\mathrm{E}[R_{j,T+1}|R_{[j-1],T+1}, \Phi_T]|\Phi_T\} \\ &= \mathrm{E}\{\hat{\alpha}_j + \hat{B}_j R_{[j-1],T+1}|\Phi_T\} \\ &= \hat{\alpha}_j + \hat{B}_j \tilde{E}_{[j-1]}. \end{aligned} \tag{A.34}$$

Applying (A.34) progressively for $j = 2, \ldots, J$ and making use of (60), (62), (63), and (A.31) establishes (67). The same analysis gives

$$\tilde{E}_{[j]} = \mathrm{E}\{R_{[j],T+1}|\Phi_T\} = \hat{E}_{[j]}, \quad j = 1, \ldots, J. \tag{A.35}$$

Computing the variance-covariance matrix of $R_{j,T+1}$ relies on the variance-decomposition rule,

$$\begin{aligned} \tilde{V}_j &= \mathrm{cov}\{R_{j,T+1},\ R'_{j,T+1}|\Phi_T\} \\ &= \mathrm{E}\{\mathrm{cov}[R_{j,T+1},\ R'_{j,T+1}|R_{[j-1],T+1}, \Phi_T]|\Phi_T\} \\ &\quad + \mathrm{cov}\{\mathrm{E}[R_{j,T+1}|R_{[j-1],T+1}, \Phi_T],\ \mathrm{E}[R_{j,T+1}|R_{[j-1],T+1}, \Phi_T]'|\Phi_T\}. \end{aligned} \tag{A.36}$$

From (A.23), computing the first term on the right-hand side of (A.36) requires the expectation of u_j, which can be rewritten as

$$\begin{aligned} u_j &= \hat{V}^{-1}_{[j-1],S_j}(R_{[j-1],T+1} - \hat{E}_{[j-1]})(R_{[j-1],T+1} - \hat{E}_{[j-1]})' \\ &\quad + 2\,(\hat{E}_{[j-1]} - \hat{E}_{[j-1],S_j})'\hat{V}^{-1}_{[j-1],S_j} R_{[j-1],T+1} \\ &\quad + \hat{E}'_{[j-1],S_j}\hat{V}^{-1}_{[j-1],S_j}\hat{E}_{[j-1],S_j} - \hat{E}'_{[j-1]}\hat{V}^{-1}_{[j-1],S_j}\hat{E}_{[j-1]}, \end{aligned} \tag{A.37}$$

and, using (A.35) and the definition of $\tilde{V}_{[j-1]}$ in (70),

$$\begin{aligned} \mathrm{E}\{u_j|\Phi_T\} &= \mathrm{tr}\,(\hat{V}^{-1}_{[j-1],S_j}\tilde{V}_{[j-1]}) + 2\,(\hat{E}_{[j-1]} - \hat{E}_{[j-1],S_j})'\hat{V}^{-1}_{[j-1],S_j}\hat{E}_{[j-1]} \\ &\quad + \hat{E}'_{[j-1],S_j}\hat{V}^{-1}_{[j-1],S_j}\hat{E}_{[j-1],S_j} - \hat{E}'_{[j-1]}\hat{V}^{-1}_{[j-1],S_j}\hat{E}_{[j-1]} \\ &= \mathrm{tr}\,(\hat{V}^{-1}_{[j-1],S_j}\tilde{V}_{[j-1]}) \\ &\quad + (\hat{E}_{[j-1]} - \hat{E}_{[j-1],S_j})'\hat{V}^{-1}_{[j-1],S_j}(\hat{E}_{[j-1]} - \hat{E}_{[j-1],S_j}). \end{aligned} \tag{A.38}$$

Therefore, combining (A.23), (A.21), and (A.38) gives

$$\mathrm{E}\{\mathrm{cov}[R_{j,T+1},\ R'_{j,T+1}|R_{[j-1],T+1}, \Phi_T]|\Phi_T\} = \kappa_j\hat{\Sigma}_j, \tag{A.39}$$

where κ_j is defined in (71). From (A.22), the second term on the right-hand side of (A.36) is equal to

$$\begin{aligned}\mathrm{cov}\{\mathrm{E}[R_{j,T+1}|R_{[j-1],T+1},\Phi_T],\ \mathrm{E}[R_{j,T+1}|R_{[j-1],T+1},\Phi_T]'|\Phi_T\} & \\ = \mathrm{cov}\{\hat{B}_j R_{[j-1],T+1}, R'_{[j-1],T+1}\hat{B}'_j|\Phi_T\} & \\ = \hat{B}_j\tilde{V}_{[j-1]}\hat{B}'_j. & \qquad \text{(A.40)}\end{aligned}$$

Combining (A.36), (A.39), and (A.40) gives

$$\tilde{V}_j = \kappa_j\hat{\Sigma}_j + \hat{B}_j\tilde{V}_{[j-1]}\hat{B}'_j,\quad j=2,\ldots,J. \qquad \text{(A.41)}$$

The covariance between $R_{j,T+1}$ and $R_{[j-1],T+1}$ is computed as

$$\begin{aligned}\mathrm{cov}\{R_{[j-1],T+1},\ R'_{j,T+1}|\Phi_T\} &= \mathrm{cov}\{R_{[j-1],T+1},\ \mathrm{E}[R_{j,T+1}|R_{[j-1],T+1},\Phi_T]'|\Phi_T\} \\ &= \mathrm{cov}\{R_{[j-1],T+1},\ R'_{[j-1],T+1}\hat{B}'_j|\Phi_T\} \\ &= \tilde{V}_{[j-1]}\hat{B}'_j,\quad j=2,\ldots,J, \qquad \text{(A.42)}\end{aligned}$$

using (A.22). Finally, combining (A.41) and (A.42) gives the result in (70).

Table 1

Parameter Estimates and Tangent-Portfolio Weights

The three return series are for (i) Standard & Poor's composite index (USA), (ii) Morgan Stanley Capital International's index for Europe, Australia, and the Far East (EAFE), and (iii) the International Finance Corporation's composite index for emerging markets (EMERGE). All returns are monthly U.S.-Dollar returns in excess of the one-month T-bill rate. The combined sample (panel A) consists of monthly returns from 1/1970–12/1995 for USA and EAFE and from 1/1985–12/1995 for EMERGE. The truncated sample (panel B) consists of monthly returns from 1/1985–12/1995 for USA, EAFE, and EMERGE.

Index	Mean	Standard Deviation	Correlations EAFE	Correlations EMERGE	Tangent Portfolio
A. Combined Sample					
		Bayesian Predictive Pdf			
USA	0.48	4.47	0.480	0.314	0.301
EAFE	0.59	5.04		0.286	0.413
EMERGE	0.71	6.70			0.286
		Maximum Likelihood			
USA	0.48	4.43	0.480	0.318	0.297
EAFE	0.59	4.99		0.290	0.410
EMERGE	0.71	6.56			0.293
B. Truncated Sample					
		Bayesian Predictive Pdf			
USA	0.89	4.35	0.429	0.306	0.528
EAFE	1.02	5.56		0.290	0.303
EMERGE	0.95	6.71			0.169
		Maximum Likelihood			
USA	0.89	4.25	0.429	0.306	0.528
EAFE	1.02	5.43		0.290	0.303
EMERGE	0.95	6.55			0.169

Table 2
Combination of Tangent Portfolio and Riskless Asset

The three return series are for (i) Standard & Poor's composite index (USA), (ii) Morgan Stanley Capital International's index for Europe, Australia, and the Far East (EAFE), and (iii) the International Finance Corporation's composite index for emerging markets (EMERGE). All returns are monthly U.S.-Dollar returns in excess of the one-month T-bill rate. The "combined sample" consists of monthly returns from 1/1970–12/1995 for USA and EAFE and from 1/1985–12/1995 for EMERGE. The "truncated sample" consists of monthly returns from 1/1985–12/1995 for USA, EAFE, and EMERGE. The parameter A denotes the investor's (approximate) coefficient of relative risk aversion.

	Combined Sample		Truncated Sample	
	Bayesian Pred. Pdf	Maximum Likelihood	Bayesian Pred. Pdf	Maximum Likelihood
A. Maximum Sharpe ratio computed using the sample and method as indicated:				
	.146	.148	.240	.245
B. Sharpe ratio of the tangent portfolio, where the Sharpe ratio is computed using the combined sample and the Bayesian predictive pdf, but the tangent portfolio is constructed using the sample and method as indicated:				
	.146	.146	.141	.141
C. Tangent portfolio proportion (%) in the overall portfolio using the sample and method as indicated:				
$A = 1$	361	369	610	639
$A = 3$	120	123	203	213
$A = 5$	72	74	122	128
D. Monthly certainty-equivalent loss (basis points) associated with the overall portfolio, where the loss is computed using the combined sample and Bayesian predictive pdf, but the overall portfolio is constructed using the sample and method as indicated:				
$A = 1$	0	0.08	55.82	67.55
$A = 3$	0	0.03	18.61	22.52
$A = 5$	0	0.01	11.17	13.51

Table 3

Estimated Standard Deviations and Weights in the Minimum-Variance Portfolio for 22 Emerging Markets

The samples and estimation methods are denoted as follows;

I Combined sample (1/89–12/95), Bayesian predictive pdf.

II Combined sample (1/89–12/95), maximum likelihood

III Truncated sample (11/93–12/95), Bayesian predictive pdf.

IV Truncated sample (11/93–12/95), maximum likelihood

V Single-series samples (using the data available for each series), maximum likelihood

The data consist of monthly returns on each country's "investable" equity portfolio, as constructed by the International Finance Corporation. All returns are U.S.-Dollar returns in excess of the one-month U.S. Treasury-bill rate.

			Standard Deviation (% per month)					Minimum-Variance Weights (×100)		
Country	First Month	No. Obs.	I	II	III	IV	V	I	II	III&IV
Argentina	1/89	84	34.8	29.2	38.9	10.6	29.2	0	2	5
Brazil	1/89	84	27.5	23.1	48.9	13.3	23.1	6	26	21
Chile	1/89	84	9.3	7.8	29.4	8.0	7.8	13	-4	-16
Greece	1/89	84	15.6	13.1	20.4	5.5	13.1	-5	-23	-32
Jordan	1/89	84	6.7	5.6	14.5	3.9	5.6	45	68	69
Malaysia	1/89	84	8.9	7.5	33.9	9.2	7.5	24	66	66
Mexico	1/89	84	12.3	10.4	47.7	13.0	10.4	8	18	13
Philippines	1/89	84	12.8	10.8	39.0	10.6	10.8	18	82	67
Portugal	1/89	84	8.3	7.0	19.0	5.2	7.0	21	17	25
Thailand	1/89	84	10.8	9.1	33.2	9.0	9.1	-10	-42	-56
Turkey	9/89	76	23.8	19.5	59.8	16.3	19.8	-0	-1	3
Venezuela	2/90	71	25.2	20.0	59.2	16.1	19.6	4	14	10
Indonesia	10/90	63	13.8	11.2	31.7	8.6	9.6	-11	-21	-2
Taiwan	2/91	59	15.2	11.7	44.4	12.1	11.1	-6	-35	-32
Colombia	3/91	58	16.6	12.7	31.8	8.7	11.4	14	46	50
Pakistan	4/91	57	16.1	11.8	36.5	9.9	11.6	-4	-16	-16
Korea	2/92	47	12.5	8.5	22.7	6.2	7.2	22	73	72
India	12/92	37	16.0	11.2	30.0	8.2	8.1	-14	-53	-44
Peru	10/93	27	23.4	14.2	44.0	12.0	11.9	-10	-41	-34
Sri Lanka	10/93	27	26.9	19.9	35.2	9.6	9.8	4	15	4
China	11/93	26	21.3	14.9	38.6	10.5	10.5	-8	-42	-33
Zimbabwe	11/93	26	25.2	19.0	34.1	9.3	9.3	-10	-49	-38

Table 4

Correlations Among 22 Emerging Markets

The upper right portion gives correlations (× 100) based on the combined-sample Bayesian predictive distribution. The lower left portion gives those values minus the maximum-likelihood estimates based on the jointly available history for a given pair (so the length of the joint history is equal to that of the shorter of the two countries). The data consist of monthly returns on each country's "investable" equity portfolio, as constructed by the International Finance Corporation. All returns are U.S.-Dollar returns in excess of the one-month U.S. Treasury-bill rate. The "combined sample" uses all available returns through 12/95, where the data for 10 countries begin in 1/89 but the data for 11 other countries begin at various later dates.

	Arg	Bra	Chi	Gre	Jor	Mal	Mex	Phi	Prt	Tha	Tur	Ven	Ido	Tai	Col	Pak	Kor	Ind	Per	Sri	Cin	Zim
Argentina		-11	5	12	-17	-4	32	7	15	13	6	17	-34	-6	-6	13	6	13	11	66	28	-8
Brazil			24	28	-6	3	20	13	27	4	9	-9	33	25	35	-7	-29	36	11	-5	-10	13
Chile				10	6	16	20	23	16	28	-5	-15	23	16	-5	5	1	26	19	-1	-12	-18
Greece					9	5	3	17	46	13	36	10	40	27	38	-3	0	58	-6	18	16	-31
Jordan						16	-3	15	-1	15	10	1	27	4	1	18	5	-4	22	-21	3	4
Malaysia							29	52	21	64	22	-10	44	35	-5	21	6	6	2	-10	42	3
Mexico								29	5	30	-4	-1	14	14	3	19	17	12	26	20	27	11
Philippines									19	56	1	-8	54	49	24	26	-20	15	14	-13	16	21
Portugal										17	23	-2	35	22	16	4	-13	34	-18	0	17	-13
Thailand											17	-9	42	36	5	24	5	14	-1	3	55	-28
Turkey	2	-3	-1	-4	0	-3	1	0	-3	-2		6	24	15	2	4	16	30	-12	19	30	-24
Venezuela	14	-4	0	3	0	1	4	1	3	2	4		-16	-14	27	0	33	15	17	45	20	39
Indonesia	-26	16	-3	8	11	-2	1	0	3	-8	-3	-18		39	26	4	-11	33	-15	-29	22	-7
Taiwan	-7	15	0	8	3	-6	-1	1	6	-5	-2	-15	0		17	5	8	9	15	-18	0	-13
Colombia	-6	15	4	23	0	-9	1	3	3	-3	-3	-7	6	6		31	-2	53	-10	35	22	39
Pakistan	7	-8	-5	-3	5	-3	-1	1	1	-3	-3	-7	-7	-4	-8		1	27	4	31	21	-3
Korea	3	-27	-17	-3	-2	-11	-17	-22	-5	-13	0	13	-16	-15	-5	-9		4	20	31	39	7
India	5	15	-30	37	-18	-12	-18	-4	0	-11	16	-3	-2	4	23	-8	-18		-41	37	32	-10
Peru	-8	-14	-34	-19	-3	-28	-47	-27	-42	-39	-8	-3	-36	-9	-20	-17	-18	-50		4	-17	10
Sri Lanka	53	-22	-43	5	-22	-24	-22	-27	-12	-14	4	11	-48	-33	-6	-11	-6	-23	-24		43	8
China	12	-14	-49	10	-7	-54	-20	-56	5	-28	0	2	-46	-57	25	-15	2	-5	-36	23		-6
Zimbabwe	-19	0	-33	-26	-10	-10	-19	-1	-11	-33	9	7	-10	-5	10	-18	10	-10	-29	-25	-10	

Table 5

Minimum Standard Deviation of a Portfolio Combining 22 Emerging Markets

The data consist of monthly returns on each country's "investable" equity portfolio, as constructed by the International Finance Corporation. All returns are U.S.-Dollar returns in excess of the one-month U.S. Treasury-bill rate. The "combined sample" uses all available returns through 12/95, where the data for 10 countries begin in 1/89 but the data for 11 other countries begin at various later dates. The "truncated sample" consists of monthly returns on all 22 countries for the 26-month period from 11/93 through 12/95.

	Combined Sample		Truncated Sample	
	Bayesian Pred. Pdf	Maximum Likelihood	Bayesian Pred. Pdf	Maximum Likelihood
A. Minimum monthly standard deviation (%) computed using the sample and method as indicated:				
	3.80	2.31	7.48	2.04
B. Monthly standard deviation (%) of the minimum-variance portfolio, where the standard deviation is computed using the combined sample and the Bayesian predictive pdf., but the weights in the minimum-variance portfolio are constructed using the sample and method as indicated:				
	3.80	6.09	6.63	6.63

Table 6

Performance in Repeated Samples: Mean-Variance Optimization

All values are true monthly certainty-equivalent losses (basis points) for portfolios constructed using the sample and method as indicated. Relative risk aversion (A) is set equal to 3. Certainty equivalents are computed based on the true moments of the multivariate normal distribution used to generate the 5000 hypothetical samples of monthly returns. The number of assets, sample size, and starting dates correspond to those used in example 1, and the true moments are set equal to the combined-sample maximum-likelihood estimates from that example (table 1).

	Combined Sample		Truncated Sample	
	Bayesian Pred. Pdf	Maximum Likelihood	Bayesian Pred. Pdf	Maximum Likelihood
A. Certainty-equivalent loss relative to the true optimal portfolio:				
mean	24.20	25.93	39.70	43.59
std. dev.	22.88	24.86	34.98	38.61
10th percentile	4.29	4.48	7.52	8.08
20	7.47	7.88	12.77	13.96
30	10.44	11.15	17.81	19.37
40	13.83	14.62	23.62	25.61
50	17.44	18.62	30.06	32.83
60	22.17	23.44	38.17	41.79
70	27.93	29.90	47.02	52.05
80	36.13	38.50	60.21	66.45
90	52.04	56.10	83.71	91.16
B. Certainty-equivalent loss relative to the Bayesian combined-sample optimal portfolio:				
mean	0	1.73	15.50	19.40
std. dev.	0	2.30	25.45	28.33
10th percentile	—	-0.11	-6.37	-4.71
20	—	0.15	-1.39	0.08
30	—	0.42	1.70	3.58
40	—	0.70	4.98	7.52
50	—	1.03	8.99	11.97
60	—	1.43	13.80	17.36
70	—	1.92	19.98	23.96
80	—	2.75	29.41	34.46
90	—	4.45	45.49	52.81

Table 7

Performance in Repeated Samples: Variance Minimization

All values are differences in true monthly standard deviations (in %), computed based on the covariance matrix of the multivariate normal distribution used to generate the 5000 hypothetical samples of monthly returns. The weights in the minimum-variance portfolios are computed using the sample and method as indicated. The number of assets, sample size, and starting dates correspond to those used in example 2 (cf. table 3), and the true moments of returns are set equal to the combined-sample maximum-likelihood estimates from that example.

	Combined-Sample, Bayesian Predictive Pdf	Combined-Sample, Maximum Likelihood	Truncated-Sample, Bayesian Pred. Pdf & Maximum Likelihood
A. Standard deviation for the minimum-variance portfolio, constructed using the sample and method as indicated, minus the true global minimum standard deviation:			
mean	2.91	4.23	3.77
std. dev.	2.02	2.92	2.39
10th percentile	1.16	1.73	1.62
20	1.46	2.20	2.04
30	1.76	2.61	2.43
40	2.06	3.02	2.82
50	2.39	3.52	3.22
60	2.79	4.04	3.65
70	3.26	4.70	4.21
80	3.96	5.69	5.03
90	5.21	7.42	6.49
B. Standard deviation for the minimum-variance portfolio, constructed using the sample and method as indicated, minus the standard deviation of the minimum-variance portfolio constructed using the combined-sample Bayesian predictive pdf:			
mean	0	1.31	0.86
std. dev.	0	1.38	1.06
10th percentile	—	0.26	0.02
20	—	0.42	0.18
30	—	0.57	0.31
40	—	0.73	0.46
50	—	0.92	0.62
60	—	1.16	0.81
70	—	1.47	1.05
80	—	1.93	1.40
90	—	2.74	2.02

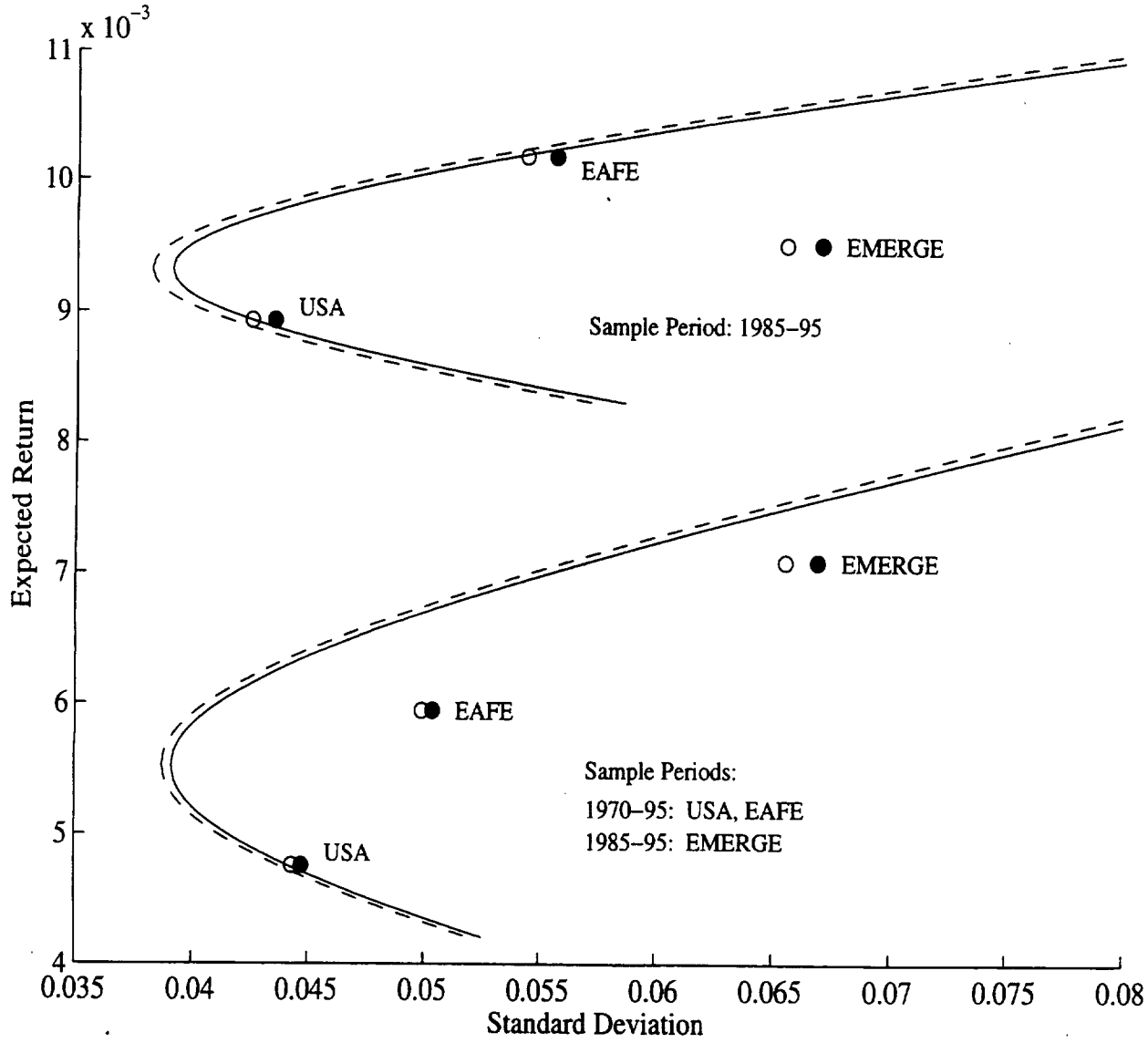

Figure 1. Minimum-Standard-Deviation Boundaries for Three Indices. The solid curves and dots are based on the Bayesian predictive pdf, whereas the dashed curves and circles are maximum-likelihood estimates. The three return series are for (i) Standard & Poor's composite index (USA), (ii) Morgan Stanley Capital International's index for Europe, Australia, and the Far East (EAFE), and (iii) the International Finance Corporation's composite index for emerging markets (EMERGE). All returns are monthly U.S.-Dollar returns in excess of the one-month U.S. Treasury-bill rate.

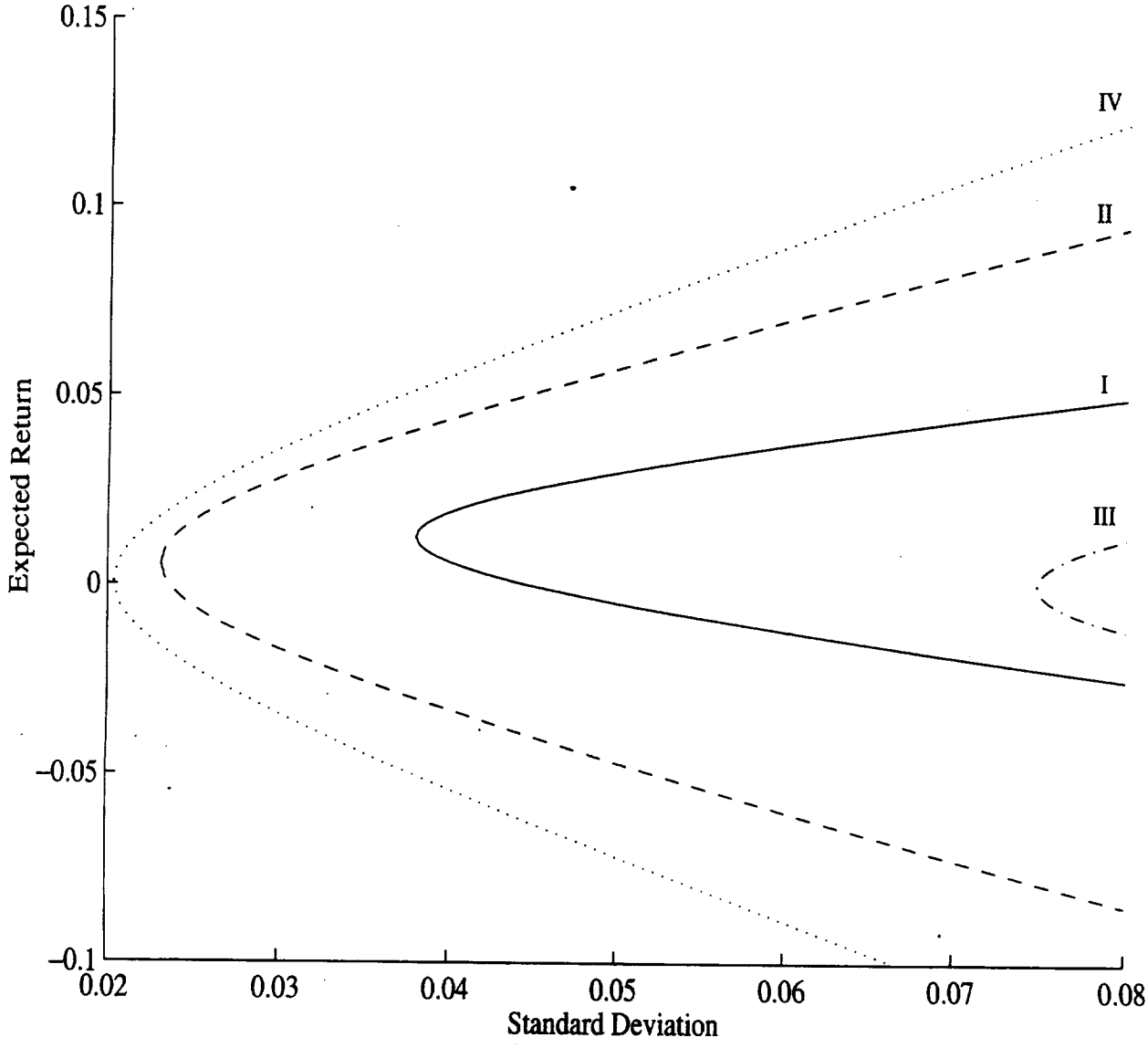

Figure 2. Minimum-Standard-Deviation Boundaries for 22 Emerging-Market Country Indices. The data consist of monthly returns on each country's "investable" equity portfolio, as constructed by the International Finance Corporation. All returns are monthly U.S.-Dollar returns in excess of the one-month U.S. Treasury-bill rate. The boundaries are estimated using four methods:

I Combined sample (1/89–12/95), Bayesian predictive pdf
II Combined sample (1/89–12/95), maximum likelihood
III Truncated sample (11/93–12/95), Bayesian predictive pdf
IV Truncated sample (11/93–12/95), maximum likelihood

References

Anderson, T.W., 1957, "Maximum Likelihood Estimates for a Multivariate Normal Distribution When Some Observations are Missing," *Journal of the American Statistical Association* 52, 200–203.

Anderson, T.W., 1984, *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York.

Barry, Christopher B. and Stephen J. Brown, 1985, "Differential Information and Security Market Equilibrium," *Journal of Financial and Quantitative Analysis* 20, 407–422.

Berger, James O., 1985, *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York.

Box, George E.P, and George C. Tiao, 1973, *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, MA.

Brown, Stephen J., 1979, "Optimal Portfolio Choice Under Uncertainty: A Bayesian Approach," in Vijay S. Bawa, Stephen J. Brown, and Roger W. Klein, Eds.: *Estimation Risk and Optimal Portfolio Choice*, North-Holland, Amsterdam.

Brown, Stephen J., William N. Goetzmann, and Stephen A. Ross, 1995, "Survival," *Journal of Finance* 50, 853–873.

Frost, Peter A. and James E. Savarino, 1986, "An Empirical Bayes Approach to Efficient Portfolio Selection," *Journal of Financial and Quantitative Analysis* 21, 293–305.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, "A Test of the Efficiency of a Given Portfolio," *Econometrica* 57, 1121–1152.

Goetzmann, William N., and Philippe Jorion, 1996, "Re-emerging Markets," working paper, Yale University and University of California at Irvine.

Grauer, Robert R., and Nils H. Hakansson, 1993, "On the Use of Mean-Variance and Quadratic Approximations in Implementing Dynamic Investment Strategies: A Comparison of Returns and Investment Policies," *Management Science* 39, 856–871.

Hansen, Lars P., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1029–1054.

Harvey, Campbell R., 1995, "Predictable Risk and Returns in Emerging Markets," *Review of Financial Studies* 8, 773–816.

Harvey, Campbell R., and Goufu Zhou, 1990, "Bayesian Inference in Asset Pricing Tests," *Journal of Financial Economics* 26, 221–254.

Ingersoll, Jonathan E., Jr., 1987, *Theory of Financial Decision Making*, Rowman and Littlefield, Savage, MD.

International Finance Corporation, 1993, *IFC Index Methodology*, Washington, D.C.

Jorion, Philippe, 1986, "Bayes-Stein Estimation for Portfolio Analysis," *Journal of Financial and Quantitative Analysis* 21, 279–292.

Kandel, Shmuel, Robert McCulloch, and Robert F. Stambaugh, 1995, "Bayesian Inference and Portfolio Efficiency," *Review of Financial Studies* 8, 1–53.

Klein, Roger W. and Vijay S. Bawa, 1976, "The Effect of Estimation Risk on Optimal Portfolio Choice," *Journal of Financial Economics* 3, 215–231.

Klein, Roger W. and Vijay S. Bawa, 1977, "The Effect of Limited Information and Estimation Risk on Optimal Portfolio Diversification," *Journal of Financial Economics* 89–111.

Little, Roderick J.A. and Donald B. Rubin, 1987, *Statistical Analysis with Missing Data*, John Wiley & Sons, New York.

Shanken, Jay, 1987, "A Bayesian Approach to Testing Portfolio Efficiency," *Journal of Financial Economics* 19, 195–215.

Tanner, Martin A., 1993, *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, Springer-Verlag, New York.

Zellner, Arnold, 1971, *An Introduction to Bayesian Inference in Econometrics*, John Wiley and Sons, New York.

Zellner, Arnold and V. Karuppan Chetty, 1965, "Prediction and Decision Problems in Regression Models from the Bayesian Point of View," *Journal of the American Statistical Association* 60, 608–616.