Bayesian Portfolio Selection: An Empirical Analysis of the S&P 500 Index 1970–1996

Nicholas G. Polson & Bernard V. Tew

Bayesian Portfolio Selection: An Empirical Analysis of the S&P 500 Index 1970–1996

Nicholas G. POLSON

Graduate School of Business, University of Chicago, Chicago, IL   60637 (ngp@gsbngp.uchicago.edu)

Bernard V. TEW

QED Investments, L.L.C., Lexington, KY   40510

In this article we present a technique for implementing large-scale optimal portfolio selection. We use high-frequency daily data to capture valuable statistical information in asset returns. We describe several statistical issues involved in quantitative approaches to portfolio selection. Our methodology applies to large-scale portfolio-selection problems in which the number of possible holdings is large relative to the estimation period provided by historical data. We illustrate our approach on an equity database that consists of stocks from the Standard and Poor's index, and we compare our portfolios to this benchmark index. Our methodology differs from the usual quadratic programming approach to portfolio selection in three ways: (1) We employ informative priors on the expected returns and variance–covariance matrices, (2) we use daily data for estimation purposes, with upper and lower holding limits for individual securities, and (3) we use a dynamic asset-allocation approach that is based on reestimating and then rebalancing the portfolio weights on a prespecified time window. The key inputs to the optimization process are the predictive distributions of expected returns and the predictive variance–covariance matrix. We describe the statistical issues involved in modeling these inputs for high-dimensional portfolio problems in which our data frequency is daily. In our application, we find that our optimal portfolio outperforms the underlying benchmark.

KEY WORDS: Asset allocation; Hierarchical models; Market efficiency; S&P 500 index; Volatility.

Optimal portfolio selection has long been of interest to both academics and practitioners. The major theoretical work showed that a quadratic programming procedure could be used to generate the optimal portfolio (Markowitz 1959). This theoretical work, however, assumed a known variance–covariance matrix. In practice, implementation has been hampered by an inability to provide a variance–covariance matrix yielding reasonable portfolios particularly for portfolios with more than 30 securities (e.g., see Michaud 1989). This article solves the implementation problem in higher dimensions. We demonstrate our technique on the equities included in the Standard and Poor's (S&P 500) index. This index comprises a value-weighted portfolio of U.S. equities. Pension funds in 1996 had approximately $400 billion managed on an index basis. The largest publicly traded mutual fund is the Vanguard Index portfolio with net assets over $80 billion. High-frequency return data, here on a daily basis, contains valuable statistical information concerning the estimation of the variance–covariance structure of future returns. Such information is lost if returns are compounded to a weekly or monthly level. The use of high-frequency daily data together with the appropriate statistical methodology results in our portfolio approach outperforming the benchmark S&P 500 index by 200 basis points (bp's) on an annual basis. Our portfolio strategy is rebalanced on a biannual window to reduce the impact of transaction costs.

Using high-frequency data requires the use of statistical models that allow for heavy-tailed distributions and possibly time-varying parameters to capture the relevant asset-return information. We assume that the standard objective of a portfolio manager is to minimize risk for a given level of expected return. Statistically, we address issues of how a manager should model the distribution of future returns and how these inputs will affect the optimization process. Our application improves on existing smaller-portfolio implementations by using nonnormal daily returns. This significantly reduces our reliance on the length of historical data needed and provides a methodology that is more appropriate when the underlying returns are nonstationary.

Several features distinguish our approach from the standard approach—(1) the use of daily return data (to incorporate the information about heavy-tailed returns), (2) a hierarchical dynamic model (to allow for possible nonstationarities), and (3) the use of informative priors and portfolio constraints. We also account for parameter and model uncertainty by computing the predictive variance–covariance matrix. We use a hierarchical, dynamic model to allow for possible nonstationarities along with informative priors and portfolio constraints. Finally, we provide new methodology for computing the predictive variance–covariance matrix when some assets have shorter historical return sequences than the rest of the securities. This is likely to occur when mergers or spin-off affect the composition of the index—for example, the Lucent Technologies spin-off from AT&T. Several works have studied the price effects of changes in

the index composition; see, for example, Harris and Gurel (1980) and Shleifer (1986).

The weighting of stocks in the S&P 500 index is, by definition, determined by the individual securities' market capitalization. In contrast, we assume that the investor has the freedom to choose the weights in an optimal way while being constrained to invest in those stocks. Capital asset-pricing model (CAPM) theory suggests that the value-weighted portfolio is mean-variance efficient; therefore, it provides a useful benchmark for our methodology. To assess clearly the outperformance of our strategy, we compute the posterior distributions of the parameters in a CAPM model of equilibrium.

Our work builds on an existing literature, applying Bayesian analysis to asset allocation and portfolio choice. Early theoretical work such as that of Zellner and Chetty (1965) and Klein and Bawa (1976), examined optimal asset allocation. More recently, our work is related to that of Black and Litterman (1993) and Putnam and Quintana (1994), who implemented Bayesian portfolio rules to problems with fixed-income securities and futures contracts in currency and equity index markets, respectively. As in this article, they used the predictive variance–covariance matrix, but their methodology has so far been limited to smaller portfolio sizes on the order of 20–30 securities.

We develop several statistical models that use historical return data to help the investor determine the inputs to solve the portfolio-optimization process. The use of a dynamic model ameliorates the stationarity assumption because the portfolio is rebalanced periodically. The choice of the updating window, or equivalently the frequency of rebalancing, is clearly related to the degree of stationarity one is willing to assume in the distribution of future returns. Under stationarity, one would be willing to use a long window and a buy-and-hold (low-turnover) portfolio strategy. In a highly nonstationary environment, one might wish to update the model frequently and rebalance the portfolio accordingly. In our empirical study, we prespecify a six-month window for reoptimizing the portfolio weights, and we use a rolling window of days for updating our distributions. The predictive minimum-variance portfolio provides a useful portfolio to study because it only depends on the predictive variance–covariance matrix. This avoids some of the issues involved with expected-return estimation. Of course, moving up the efficient frontier to the diversified mean-variance portfolio would lead to improved performance, but it also requires an accurate model of future expected returns.

Although the mean-variance quadratic programming framework does not depend on the assumption of multivariate normality, the sample variance–covariance matrix typically performs poorly under nonnormal returns. It is also well known that the S&P 500 index's daily returns exhibit nonnormality (e.g., see Jacquier, Polson, and Rossi 1994, 1995). One of the main problems with mean-variance models is that the mapping between variance-covariance/expected-return inputs and optimal portfolio weights is highly nonlinear and can be very sensitive to small changes in the views of the manager (e.g., see, Jobson

and Korkie 1980). Chopra and Ziemba (1993) concluded that estimation error in the expected-return inputs is typically 10 times more damaging than similar errors in variances estimation and more than 100 times more damaging than errors in covariance estimation.

In asset-allocation problems, there are several possibilities for modeling prior information. One of the advantages of the hierarchical approach is that it allows the investor to reflect financial information or private views about the future return-generating process rather than just the historical returns. This framework suggests that portfolio optimization should be seen as an iterative approach in which individual views are combined with market equilibrium to form portfolios, which then, based on their composition or performance, lead the investor to update their initial opinions. Black and Litterman (1993) provided an example in which subjective prior information can be used to affect the asset-allocation procedure.

This article is organized as follows. Section 1 develops the portfolio-optimization framework with parameter uncertainty. Section 2 describes several different approaches to modeling the multivariate distribution of returns. We focus on a three-stage hierarchical model because this class of models is both flexible and implementable in higher dimensions. The predictive variance–covariance matrix is developed for several different hierarchical models. In particular, we develop a methodology that handles the case in which some of the securities have data histories much less than the rest of the securities. Section 3 presents an empirical study implementing our methodology on the portfolio defined by the S&P 500 from January 1970 to December 1996.

1. PORTFOLIO ALLOCATION

In this section, we review portfolio-allocation problems and relax the standard assumptions of known parameters and stationarity. We begin by reviewing the standard framework and its shortcomings from assuming parameter certainty. Then we describe the predictive Bayesian approach, using hierarchical models that incorporate parameter uncertainty and nonstationarity.

Consider a model evolving for T time periods. Let k denote the number of assets and let $R_1 = (R_{11}, \ldots, R_{1k})$ be the random return vector representing the rates of return on the assets from period 1. Similarly, define R_i for $i = 2, \ldots, T$. A typical assumption is that each vector R_i is identically distributed with the random vector R having mean $\mu = E(R) = (\mu_1, \ldots, \mu_k)$ and covariance matrix $\Sigma = \text{cov}(R) = E((R - \mu)(R - \mu)')$. The standard approach is based on using historical data to estimate the parameters μ and Σ. A portfolio at time t is denoted by a vector $\mathbf{x} = (x_1, \ldots, x_k)$, where x_i is the proportion of wealth invested in asset i, where $1 \leq i \leq k$. The usual portfolio constraint gives $\iota'\mathbf{x} = 1$ with $\iota = (1, \ldots, 1)'$, a vector of ones (suppressing the time subscript for clarity). For the purpose of simplicity, let us assume that the variance–covariance matrix is positive definite.

The standard approach assumes that the investor will adopt a strategy that, for some fixed mean rate of return

on their portfolio $P, \mu_P = E(\mathbf{x}^T R) = \mathbf{x}^T \mu$, minimizes the variance of the return over admissible portfolios $\mathbf{x}$. The variance of P is given by $\sigma_P^2 = \mathbf{x}^T \Sigma \mathbf{x}$. Subject to a set of linear constraints, the *efficient frontier* is the set of portfolios that have maximal expected return given an upper bound on the variance. It can be shown that the efficient frontier is the solution $\mathbf{x}_\lambda$ that minimizes $\frac{1}{2}\mathbf{x}^T \Sigma \mathbf{x} - \lambda \mathbf{x}^T \mu$ over the constraint set (Markowitz 1959).

The problem for the investor then reduces to solving the quadratic programming problem:

$$\min_{\mathbf{x}} \frac{1}{2} \mathbf{x}^T \Sigma \mathbf{x} \quad \text{subject to} \begin{cases} \mathbf{x}^T \iota = 1 \\ \mathbf{x}^T \mu = \mu_P \end{cases}.$$

For simplicity, consider the unconstrained problem in which we can identify the *mean/variance efficient portfolio*,

$$\mathbf{x}_{\text{EV}} = \frac{1}{\iota^T \Sigma^{-1} \mu} \Sigma^{-1} \mu,$$

with expected return $\mu_{\text{EV}} = \mu^T \Sigma^{-1} \mu / \iota^T \Sigma^{-1} \mu$. We can also define the *minimum variance portfolio* $\mathbf{x}_{\text{MV}}(\Sigma)$ by

$$\mathbf{x}_{\text{MV}}(\Sigma) = \frac{1}{\iota^T \Sigma^{-1} \iota} \Sigma^{-1} \iota,$$

which has expected return $\mu_{\text{MV}} = \iota^T \Sigma^{-1} \mu / \iota^T \Sigma^{-1} \iota$. The global minimum-variance portfolio just depends on the variance–covariance matrix Σ and so, from a statistical viewpoint, becomes a good portfolio to study as we change the input Σ.

As discussed by Perold (1988) and Chopra and Ziemba (1993), implementation of portfolio choice in higher dimensions tends to result in extreme weights on securities. One strategy to approach this issue is to introduce constraints in the optimization problem. We introduce upper and lower constraints in the optimization problem by letting $c_i(x_i) = -\infty$ if $x_i < l_i$ or $x_i > u_i$ and consider the problem of

$$\min_{\mathbf{x}} \frac{1}{2} \mathbf{x}^T \Sigma \mathbf{x} - \mathbf{x}^T \mu - \lambda c(\mathbf{x}),$$

where $c(\mathbf{x}) = \Sigma_{i=1}^k c_i(x_i)$. In the indexing problem, we will typically choose $c_i(x_i) = c$ on $l_i < x_i < u_i$ and $l_i = 0 \ \forall i$ (reflecting a no-short-sales constraint) and $u_i = u_0$, a predetermined constant upper bound. The higher the level of u_0 the more aggressive the portfolio in the sense of the few numbers of securities held and the higher tracking error of the portfolio. Other choices could depend on the benchmark index weights or the individual volatilities of the securities.

1.1 Bayesian Portfolio Selection Incorporating Parameter Uncertainty

There are two well-known problems with the standard approach. First, misspecification of the input parameters (μ, Σ) leads to poorly performing portfolios, particularly the predictive expected return vector (see Chopra and Ziemba 1993; Frost and Savarino 1988). Second, following Zellner and Chetty (1965) and Bawa and Klein (1976), the optimization should take account for parameter uncertainty. In particular, the rational investor would maximize expected utility over the predictive distribution of future

returns. The predictive mean–variance analysis developed here is likely to provide only an approximation to the general expected utility problem. If the predictive $p(R_{t+1}|D_t)$ lies in the class of elliptical distributions, then our approach has an expected utility justification. Investors will also have to incorporate dynamic aspects to cope with the possibility of nonstationarity of returns. In all cases, the required inputs will be predictive means and the predictive variance–covariance matrix of returns.

Suppose that we have observed data up to time t and we are solving the optimization (myopically) for time interval $(t, t+1)$. Let D_t denote the historical return data observed up to time t. In the Bayesian framework, we have a posterior distribution of future returns given by $p(R_{t+1}|D_t)$. This leads to a quadratic optimization given by

$$\min_{\mathbf{x}} \frac{1}{2} \mathbf{x}^T \text{var}(R_{t+1}|D_t) \mathbf{x}$$

$$\text{subject to} \begin{cases} \mathbf{x}^T \iota = 1 \\ \mathbf{x}^T E(R_{t+1}|D_t) = \mu_P \end{cases},$$

where $\Sigma_P = \text{var}(R_{t+1}|D_t)$ is the *predictive* variance–covariance matrix of the future returns and the forecast of future expected returns is $\mu_{t+1} = E(R_{t+1}|D_t)$. We can still compute the minimum-variance portfolio $\mathbf{x}(\Sigma_P)$ as a function of the predictive variance–covariance matrix.

The following proposition, a standard result in hierarchical modeling, provides the basis for computing the predictive mean and variance–covariance matrix.

1.1.1 Proposition 1. The predictive variance–covariance matrix $\Sigma_P = \text{var}(R_{t+1}|D_t)$ can be determined by

$$\Sigma_P = E(\Sigma|D_t) + \text{var}(\mu|D_t). \tag{1}$$

This is because

$$\Sigma_P = \text{var}(R_{t+1}|D_t)$$

$$= E(\text{var}(R_{t+1}|\mu, \Sigma)|D_t) + \text{var}(E(R_{t+1}|\mu, \Sigma)|D_t),$$

whereas $\text{var}(R_{t+1}|\mu, \Sigma, D_t) = \Sigma$ and $E(R_{t+1}|\mu, \Sigma, D_t) = \mu$.

2. MULTIVARIATE HIERARCHICAL MODELING

In this section, we describe a factor hierarchical approach to modeling the distribution of daily returns for stocks that possibly have different sample length. Specifically, consider a sample of N firms with $T_i, i = 1, \dots, N$, daily observations, and define r_i as the $(T_i \times 1)$ column vector of firm i's returns in excess of the risk-free rate, f_i as the $(T_i \times k)$ matrix of factor-mimicking portfolios' returns, and β_i as the $(k \times 1)$ vector of factor loadings. We will use this model in the empirical section, where we have a one-factor model in which the factor is defined by the returns on the S&P 500 index itself.

The firm's excess returns are modeled as follows: $r_i = f_i \beta_i + \varepsilon_i \ \forall i = 1, \dots, N$. The system of all N assets is written using a seemingly unrelated regression setup

$$R = FB + E, \tag{2}$$

where

$$
R = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix}, \quad F = \begin{bmatrix} f_1 & 0 & 0 & 0 \\ 0 & f_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & f_N \end{bmatrix}, \quad E = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}
$$

R is a $(\sum_{i=1}^{N} T_i \times 1)$ stacked vector of firm returns. F is a $(\sum_{i=1}^{N} T_i \times kN)$ block diagonal matrix of factor realizations, and B is a $(kN \times 1)$ vector of stacked factor loadings $\beta_1, \ldots, \beta_N$. E is a $(\Sigma_{i=1}^{N} T_i \times 1)$ stacked vector of firm residuals.

Let E be multivariate normal with mean 0 and a $(\sum_{i=1}^{N} T_i \times \sum_{i=1}^{N} T_i)$ variance–covariance matrix Σ. The likelihood function $l(E|B, \Sigma)$ is multivariate normal,

$$
l(E|B, \Sigma)
$$

$$
\propto |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2}(R - FB)'\Sigma^{-1}(R - FB) \right\}. \quad (3)
$$

The prior for B is formed using a hierarchical multivariate normal setup. See Stambaugh (1997, p. 323) for discussions regarding the use of informative priors. Each β_i is assumed to be an independent and identical draw from the following multivariate normal distribution: $\beta_i \sim \mathcal{N}(\bar{\beta}, \Delta_\beta) \; \forall i = 1, \ldots, N$, where $\bar{\beta}$ is a $(k \times 1)$ mean vector and Δ_β is a $(k \times k)$ diagonal matrix with elements $(\delta_1, \ldots, \delta_k)$. The joint posterior distribution for the parameters and hyperparameters of the model is

$$
p(B, \Sigma, \bar{\beta}, \Delta_\beta | R, F)
$$

$$
\propto l(E|B, \Sigma)p(\Sigma)p(B|\bar{\beta}, \Delta_\beta)p(\bar{\beta})p(\Delta_\beta) : \quad (4)
$$

The multivariate normal distribution for B is

$$
B|\bar{\beta}, \Delta_\beta, \Sigma, R, F \sim \mathcal{N}(b^*, (F'\Sigma^{-1}F + I_N \otimes \Delta_\beta^{-1})^{-1}), \quad (5)
$$

where

$$
b^* = (F'\Sigma^{-1}F + I_N \otimes \Delta_\beta^{-1})^{-1}
$$

$$
\times ((F'\Sigma^{-1}F)\hat{b}_{\text{gls}} + (I_N \otimes \Delta_\beta^{-1})(\iota_N \otimes \bar{\beta})), \quad (6)
$$

where ι_N is an $(N \times 1)$ vector of ones, I_N is an $(N \times N)$ identity matrix, and $\hat{b}_{\text{gls}}$ is a vector of generalized least squares regression coefficients; namely, $\hat{b}_{\text{gls}} = (F'\Sigma^{-1}F)^{-1}F'\Sigma^{-1}R$. The multivariate normal distribution for the hyperparameter vector $\bar{\beta}$ is

$$
\bar{\beta}|B, \Delta_\beta \sim \mathcal{N}\left(\frac{1}{N}(\iota_N \otimes I_k)'B, \frac{1}{N}\Delta_\beta\right), \quad (7)
$$

where I_k is a $(k \times k)$ identity matrix. The Wishart distribution for the precision matrix Δ_β^{-1} is

$$
\Delta_\beta^{-1}|B, \bar{\beta}
$$

$$
\sim \mathcal{W}(\nu_\beta + N, ((D - \bar{\beta} \otimes \iota_N')(D - \bar{\beta} \otimes \iota_N')' + \Psi)^{-1}),
$$

where Ψ is a prior precision matrix and D is a $(k \times N)$ matrix whose N columns, each of length k, are taken sequentially from the vector B.

The next section describes methods for computing predictive variance–covariance matrices given a historical set of returns.

2.1 Hierarchical Predictive Variance–Covariance Matrices

We now discuss methods for calculating the predictive variance–covariance matrix defined by (1) for several hierarchical models. One strategy for determining the variance–covariance matrix Σ_P for input into the optimization process would be based on a historical string of past returns. The estimation procedure will depend crucially on the type of distributional assumptions you are willing to make about the returns. For example, Markowitz (1959) proposed the use of semivariance estimators, which focus only on the information in the negative returns. Black and Litterman (1993) proposed a fully subjective approach that combines the investor's private views with that of market equilibrium. Finally, we propose a method based on three-stage hierarchical models. For applications of hierarchical models in smaller portfolio contexts, see Brown (1979), Frost and Savarino (1986), Jorion (1986), Dumas and Jacquillant (1990), and Putnam and Quintana (1994). In our implementation, we will also add a dynamic aspect.

First, consider using an approach based on historical returns. Suppose that we have a dataset of returns:

$$
R = \{R_1, \ldots, R_t\}
$$

$$
= \{(R_{11}, \ldots, R_{1k}), \ldots, (R_{t1}, \ldots, R_{tk})\}.
$$

Let Σ denote the covariance matrix and let D_t denote the data history observed up to time t. In our hierarchical framework, we will model the distribution of returns as a location-scale mixture of normality. That is,

$$
p(R) = \int_{\mu, \Sigma} (\det(\Sigma))^{-t/2}
$$

$$
\times \exp\left(-\frac{1}{2}\sum_{i=1}^{t}(R_i - \mu)'\Sigma^{-1}(R_i - \mu)\right)
$$

$$
\times p(\mu, \Sigma)d\mu d\Sigma. \quad (8)
$$

Under this model, the predictive distribution of returns is given by

$$
p(R_{t+1}|D_t) = \int_{\mu, \Sigma} p(R_{t+1}|\mu, \Sigma)p(\mu, \Sigma|D_t)d\mu d\Sigma,
$$

where $p(\mu, \Sigma|D_t)$ denotes the updated posterior distribution of the parameters of the model.

We now determine several predictive variance–covariance matrices for different specifications of $p(\mu, \Sigma)$.

Noninformative Prior Approach. Several authors have proposed using a purely noninformative prior approach to the problem. The following lists the associated predictive distribution.

Proposition 2. Under the prior $p(\mu, \Sigma) \propto (\det(\Sigma))^{-1/2(k+1)}$ the predictive distribution of future returns is multivariate t with

$$p(R_{t+1}|D_t)$$

$$\propto \left| (t-1)\hat{\Sigma} + \frac{t}{t+1}(R_{t+1} - \bar{R})(R_{t+1} - \bar{R})' \right|^{-1/2(k+1)},$$

where $\bar{R} = 1/t \sum_{i=1}^{t} R_i$ and $\hat{\Sigma} = 1/(t-1) \sum_{i=1}^{t} (R_i - \bar{R})(R_i - \bar{R})'$ and therefore

$$\Sigma_P = \text{var}(R_{t+1}|D_t) = \left(1 + \frac{1}{t}\right) \frac{t-1}{t-k-2} \hat{\Sigma}.$$

Because Σ_P is proportional to the standard maximum likelihood estimator $\hat{\Sigma}$, the predictive minimum-variance portfolio $x_{\text{MV}}(\Sigma_P) = x_{\text{MV}}(\Sigma)$ and there is no effect of parameter uncertainty on the portfolio rule. We find that this does not lead to superior performance in our empirical analysis of the S&P 500 data.

Informative Hierarchical Prior Approach. The mixture model described previously (8) can be thought of as a three-stage hierarchical model corresponding to a sequence of conditional distributions $p(R|\mu, \Sigma), p(\mu|\Sigma)$, and $p(\Sigma)$ for the distribution of the data, given parameters and the distribution of the parameters. A useful model introduces a scaling λ^{-1} of the prior variance–covariance matrix that can be estimated from the data and is specified by the conditional structure

$$R|\mu, \Sigma \sim \mathcal{N}(\mu, \Sigma)$$

$$\mu|\Sigma, \lambda \sim \mathcal{N}(\mu_0, \lambda^{-1}\Sigma)$$

$$\Sigma \sim \mathcal{W}^{-1}(\nu_0, \Sigma_0),$$

where $\lambda \sim Ga(\alpha_0, \beta_0)$ and $W^{-1}(\nu_0, \Sigma_0)$ denotes the inverse Wishart distribution with hyperparameters (ν_0, Σ_0). The prior hyperparameters $(\mu_0, \lambda, \nu_0, \Sigma_0)$ determine the portfolio manager's strength of belief in the prior means (μ_0, Σ_0) for the expected returns and variance–covariance matrix. See Black and Litterman (1993) for an alternative informative hierarchical approach.

Proposition 3. Under the prior $p(\mu|\lambda, \Sigma) \sim N(\mu_0, \lambda^{-1}\Sigma)$ and $p(\Sigma) \sim W^{-1}(\nu, S_0)$, the predictive variance–covariance matrix is given by

$$E(R_{t+1}|D_t) = \frac{(t\bar{R} + \lambda\mu_0)}{t + \lambda}$$

$$\Sigma_P = \text{var}(R_{t+1}|D_t)$$

$$= C(\hat{\Sigma} + S_0 + \lambda\mu_0\mu_0^T + t\bar{R}\bar{R}'$$

$$- (\lambda + t)E(R_{t+1}|D_t)E(R_{t+1}|D_t)^T),$$

where

$$C = \left(1 + \frac{1}{t+\lambda}\right) \frac{\nu_0 + t}{(\nu_0 + t - 2)(t - k - 2)}.$$

In Section 3, we will use this model in a dynamic framework in which we use a rolling-estimation window of past historical returns for estimating the predictive variance–covariance matrix Σ_P of the next biannual holding period.

The difficulties and lack of robustness of using sample estimates have been widely noted in the finance literature. We will see that the use of an informative hierarchical prior structure alleviates this difficulty. When a security has a small data history, we will also have to use an informative-prior specification to address the problem. We now describe an approach to modeling Σ_P in such contexts.

3. MODELING Σ FOR ASSETS WITH DIFFERENT HISTORICAL–DATA LENGTHS

A typical implementation problem occurs when a few assets, p say, in the sample have historical return length that is less than the rest of the assets. This is an issue in the empirical section because stocks can enter the S&P 500 periodically due to mergers and spin-offs. Clearly, in the case of spin-offs—for example, the Lucent Technology spin-off from AT&T in 1996—we have very little information concerning the historical returns on the asset. Stambaugh (1996) provided one approach to this problem. We solve the problem by providing a simple Gibbs-sampling approach that is based on developing a flexible family of priors on the variance–covariance matrix.

Let $R_{1,t}$ contain the returns on the $k - p$ assets that we have for times $t = 1, \ldots, T$ and let $R_{2,t}$ contain the returns on the other p assets that are only observed for times $t = s, \ldots, T$, where $s \geq 1$. Conditional on μ and Σ, we have

$$E\begin{pmatrix} R_{1,t} \\ R_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

$$\text{cov}\left(\begin{pmatrix} R_{1,t} \\ R_{2,t} \end{pmatrix}, (R_{1,t}, R_{2,t})\right) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Here Σ is the full $k \times k$ variance–covariance matrix (assumed positive semidefinite). The inverse Wishart distribution has density $p(\Sigma|\nu, S) \propto (\det(\Sigma))^{-(q+\nu/2)} \exp(-\frac{1}{2}\text{tr}(\nu S\Sigma^{-1}))$, where $\nu > 0$ is a known scalar degrees-of-freedom parameter and S is a known $q \times q$ positive definite matrix. If $\nu > 2$, then $E(\Sigma|\nu, S) = \nu S/(\nu - 2)$. We denote this family by $\Sigma \sim W^{-1}(\nu, S)$. One of the caveats of the inverse Wishart distribution is that it is not flexible enough to model constraints on specific elements of the matrix. Here the investor knows more about Σ_{11} than Σ_{22}, for instance, due to the differences in the length of data histories.

The likelihood function is given by

$$p(R_{1,1}, \ldots R_{1,s-1}, R_s, \ldots, R_T|\mu, \Sigma)$$

$$= p(R_{1,1}, \ldots, R_{1,s-1}|\mu_1, \Sigma_{11})p(R_s, \ldots, R_T|\mu, \Sigma).$$

Therefore,

$$L(\mu, \Sigma|D_t) \propto \det(\Sigma_{11})^{(s-1)/2} \prod_{t=1}^{s-1}$$

$$\times \exp\left(-\frac{1}{2}(R_{1,t} - \mu_1)'\Sigma_{11}^{-1}(R_{1,t} - \mu_1)\right)$$

$$\det(\Sigma)^{(T-s)/2} \prod_{t=s}^{T} \exp\left(-\frac{1}{2}(R_t - \mu)'\Sigma^{-1}(R_t - \mu)\right).$$

Table 1. Nifty 50 Stocks Weighting in the S&P 500 Index

Date	Symbol	6/96	12/89	12/79	12/69
General Electric	GE	2.800	2.485	1.640	1.569
Coca Cola	KO	2.342	1.126	.606	1.051
Exxon	XON	2.142	2.672	3.439	2.957
AT&T	T	2.030	2.090	5.197	5.948
Philip Morris	MO	1.678	1.649	.637	*****
Royal Dutch	RD	1.636	1.774	1.191	*****
Merck	MRK	1.615	1.308	.773	.906
Microsoft	MSFT	1.436	*****	*****	*****
Johnson/Johnson	JNJ	1.320	.845	.689	*****
Intel	INTC	1.262	*****	*****	*****
Proctor and Gamble	PG	1.228	1.040	.871	.993
Walmart	WMT	1.208	1.084	*****	*****
IBM	IBM	1.181	2.327	5.341	9.231
Hewlett Packard	HWP	1.105	.477	.497	*****
Pepsi	PEP	1.061	.719	*****	*****
Pfizer	PFE	.918	.491	.408	.486
Dupont	DD	.910	1.229	.837	1.101
AIG	AIG	.910	.723	*****	*****
Mobil	MOB	.906	1.093	1.659	1.040
Bristol Myers Squibb	BMY	.878	1.247	*****	.484
GTE	GTE	.849	.975	.593	.705
General Motors	GM	.848	1.086	2.079	4.399
Disney	DIS	.839	.644	*****	*****
Citicorp	CCI	.831	.400	.418	*****
BellSouth	BLS	.822	1.190	*****	*****
Motorola	MOT	.804	*****	*****	*****
Ford	F	.798	.883	.485	.640
Chervon	CHV	.794	.990	1.370	.966
Amoco	AN	.733	1.198	1.673	.758
Eli Lilly	LLY	.720	.814	*****	*****
Abbott Labs	ABT	.690	.654	*****	*****
AmerHome Products	AHP	.686	.716	.606	.793
FedNatlMortgage	FNM	.686	*****	*****	*****
McDonald's	MCD	.686	.545	*****	*****
Ameritech	AIT	.639	.782	*****	*****
Cisco Systems	CSCO	.633	*****	*****	*****
CMB	CMB	.621	*****	*****	*****
SBC	SBC	.612	.819	*****	*****
Boeing	BA	.598	.584	.462	*****
MMM	MMM	.581	.762	.838	1.331
BankAmerica	BAC	.560	*****	.577	*****
Bell Atlantic	BEL	.556	.946	*****	*****
Gillette	G	.535	*****	*****	*****
Kodak	EK	.524	.570	1.106	*****
Chrysler	C	.507	*****	*****	.367
Home Depot	HD	.497	*****	*****	*****
Colgate	COL	.489	.499	*****	*****
Wells Fargo	WFC	.478	*****	*****	*****
Nations Bank	NB	.453	*****	*****	*****
American Express	AXP	.450	.621	*****	*****

Consider the reparameterization of Σ given by $\Sigma \rightarrow (\Sigma_{11}, \psi, \Sigma_{22.1})$, where $\psi = \Sigma_{11}^{-1}\Sigma_{12}$ and $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. The family of distributions is defined by the hierarchical model

$$\Sigma_{11} \sim \mathcal{W}^{-1}(\nu_{11}, \Omega_{11})$$

$$\psi|\Sigma_{22.1} \sim \mathcal{MN}(\psi_0, \Omega_0, \Sigma_{22.1})$$

$$\Sigma_{22.1} \sim \mathcal{W}^{-1}(\nu_{22}, \Omega_{22.1}). \qquad (9)$$

Consider that a prior distribution on Σ is given by $p(\Sigma) = p(\Sigma_{11})p(\psi|\Sigma_{22.1})p(\Sigma_{22.1})$ and assume that Σ_{11} is independent of $(\psi, \Sigma_{22.1})$. Here $(\nu_{11}, \Omega_{11}, \psi_0, \Omega_0, \nu_{22}, \Omega_{22.1})$ are known hyperparameters.

The degree-of-freedom parameters $\nu_{11}, \nu_{22} > 0$ and the prior mean matrices Ω_{11}, Ω_{22} are $p \times p$ and $(k -$

$p) \times (k - p)$ positive definite matrices. The distribution $\mathcal{MN}(\psi_0, \Omega_0, \Sigma_{22.1})$ denotes the matrix normal distribution where ψ_0 is a $p \times k$ mean matrix, Ω_0 is a $p \times p$ left-variance matrix, and $\Sigma_{22.1}$ is a right-variance matrix. The matrix normal density is given by $p(\psi|\Sigma_{22.1}) \propto \exp(-\frac{1}{2}\text{tr}(\Sigma_{22.1}^{-1}(\psi - \psi_0)'\Omega_0(\psi - \psi_0))$; see Dawid (1981) and West and Harrison (1989, p. 619). The marginal distribution for ψ is a matrix t distribution.

The following proposition characterizes the posterior distribution for $\Sigma|\mu, D_t$ via the conditional distributions needed in the Gibbs sampler to simulate this distribution. This can then be used to determine the predictive variance–covariance matrix for the future returns.

Proposition 4. The posterior distribution $p(\Sigma|\mu, D_t)$ for the likelihood and prior defined previously is given by the sequence of conditional distributions

$$\Sigma_{11}|\mu, D_t \sim \mathcal{W}^{-1}\left(\nu_{11} + T, \frac{\nu_{11}}{\nu_{11} + T}\Omega_{11} + \frac{T}{\nu_{11} + T}R_{11}^T\right)$$

$$\psi|\Sigma_{22.1}, \mu, D_t \sim \mathcal{N}(\psi_{(T-s)}, \Omega_{(T-s)}, \Sigma_{22.1})$$

$$\Sigma_{22.1}|\psi, \mu, D_t \sim \mathcal{W}^{-1}\left(\nu_{22} + (T - s),\right.$$

$$\left.\frac{\nu_{22}}{\nu_{22} + (T - s)}\Omega_{22.1} + \frac{T - s}{\nu_{22} + (T - s)}(R_{22.1} + R^*(\psi))\right),$$
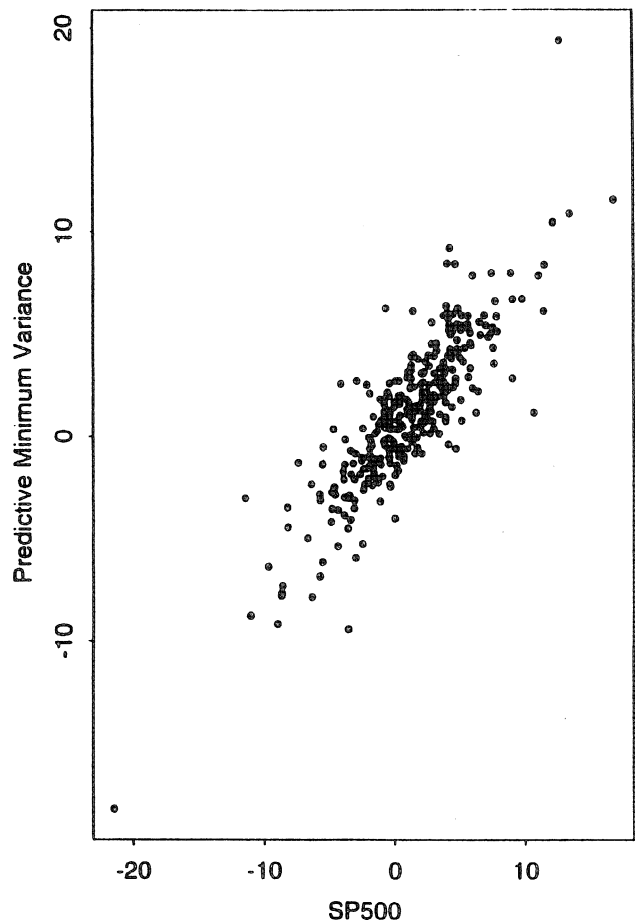

Figure 1 Predictive MinVariance Versus the S&P 500.

where the updated hyperparameters are defined by

$$\psi_{(T-s)} = ((T-s)R_{11} + \Omega_0)^{-1}(\Omega_0\psi_0 + (T-s)R_{11})$$

$$\Omega_{(T-s)} = ((T-s)R_{11} + \Omega_0)$$

$$R^*(\psi) = R_{22.1} + (\psi - R_{11}^{-1}R_{12})'R_{11}(\psi - R_{11}^{-1}R_{12}).$$

Proof. See Appendix.

Section 4 provides the empirical results of the article on an equity database of stocks from the S&P 500 index.

4. PORTFOLIO ALLOCATION IN THE S&P 500 INDEX

We consider daily return data on the S&P 500 index from the time period January 1, 1970, to December 31, 1996. By definition, the S&P 500 portfolio is a value-weighted index of the stocks involved. The key to the indexing approach is that the portfolio weights are solely determined by the current market valuation. The market-capitalization weights of the individual securities are themselves highly variable over time. Table 1 shows how the portfolio weighting has varied for the current top 50 stocks in the composition of the index.

The database of historical returns is adjusted for dividends, stock splits, rights offerings, and mergers. Although the index approach is thought of as a passive approach to investing, there are frequent typically minor changes in the index composition. Tracking such a portfolio can be costly, especially when transaction costs are included. In our backtest methodology, we recognize that, at each subsequent optimization period, companies exit or enter the universe as their holding in the S&P 500 dictates. Our portfolio is rebalanced on a biannual window, and the database is backfilled for those companies that have entered the index. The reason for a biannual window is to reduce the effect of transaction costs that here would be in the range of about 50 bp's per year. Given that the outperformance of our portfolio strategy is typically around 200 bp's, this seems a reasonable choice of rebalancing window.

The estimation method described in Section 2 is used for the companies without the necessary data history. This is of particular interest in situations in which there are additions, mergers, and spin-offs in the index. In the period 1/1/94 to 12/31/96, there have been 73 replacements to the S&P 500 index. For example, the merger of Chase and Chemical Bank on April 1, 1996, resulted in the removal of the (old)

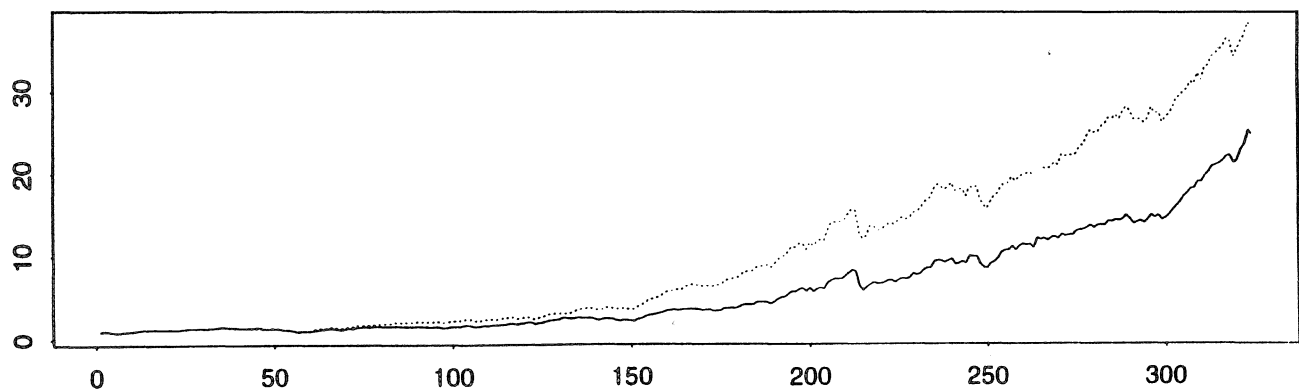## Monthly Returns
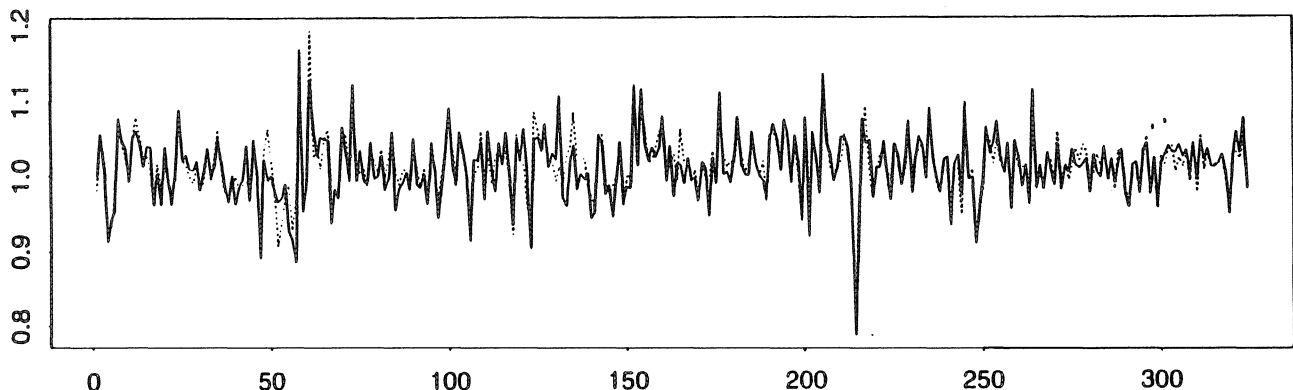
Wealth Relatives

Figure 2. Monthly Returns and Wealth Relatives for S&P 500 and Optimized Portfolio.

Chase and Chemical and the addition of the (new) Chase and a new company, WorldCom, to reproduce the stocks. After portfolios are invested, no further adjustment is made until the next rebalancing period even though the S&P 500 often experiences deletions, additions, and/or other adjustments.

To illustrate our methodology, we consider the informative-prior approach of Section 2. The model also has a dynamic aspect to how it updates the parameters. We use a biannual window for reoptimizing and rebalancing the portfolio together with an estimation window to compute the predictive variance–covariance matrix. We focus on the predictive minimum-variance optimal portfolio because it is insensitive to our model specification for expected returns. A critical line algorithm is used to implement the quadratic programming step. We use a constraint vector that prohibits short sales (so $x_i \geq 0$ for each i) and places an upper bound of 2% of assets in any one security (so $x_i \leq .02$ for each i). Constraints on portfolio weights are imposed for several reasons. Upper and lower bounds on the holdings are imposed partly because of institutional restriction (e.g., no short sales) and partly because of the sensitivity of weights to the modeling of expected returns and covariances in such high-dimensional problems. We would like our portfolio to track the returns of the index but in a manner such that the optimized portfolio lies closer to the ex post efficient frontier than the S&P 500 index. Notice that it is quite reasonable for the optimal solution to place zero (assuming this is the lower bound on portfolio weights) weight in several holdings. In fact, one of the interesting implications of the quantitative approach to asset-allocation problems is the way that holdings get added or subtracted from the optimal solution. For example, in the S&P 500 problem, it is easy to empirically show that the number of securities in the optimal solution decreases as you move down the efficient frontier to the lower volatile securities.

There is a large literature on how to measure the performance of a portfolio relative to a benchmark index. One interesting line of research has tried to develop Bayesian measures of efficiency of a portfolio. That is, can we measure how close the portfolio is to the efficient set of portfolios? Kandel, Stambaugh, and McCulloch (1996) provided several measures to address this issue. Let $R_{i,P}$ denote the observed returns on the optimized portfolio, let $R_{i,M}$ denote the returns on the market index, and let $R_{i,F}$ denote the risk-free rate in period i. A standard procedure would be to estimate the CAPM model $(R_{i,P} - R_{i,F}) = \alpha + \beta(R_{i,M} - R_{i,F}) + \varepsilon_i$, where $\varepsilon_i \sim N(0, \sigma^2)$ and $p(\alpha, \beta) \sim N(\mu_0, \Sigma_0)$, where a priori $E(\alpha) = 0, E(\beta) = 1$, and Σ_0 is relatively diffuse. The reason for assuming that the prior is centered on $\beta_0 = 1$ is that by construction the portfolio is trying to mimic the market. The prior for the intercept α is centered on $\alpha_0 = 0$ representing prior ignorance as to whether the portfolio can obtain an excess expected return or not. See Vasicek (1973) for hierarchical approaches for implementing this model. Portfolio efficiency can be assessed by the posterior for α, $p(\alpha|D_t)$, and sensitivity to market changes by $p(\beta|D_t)$.

Figures 1 and 2 show a plot of the monthly returns for the optimized predictive minimum-variance portfolio returns versus the market S&P 500 portfolio. Figure 3 shows the posterior distributions for $p(\alpha|D_t)$ and $p(\beta|D_t)$. We see that $E(\alpha|D_t) = .39$—that is, 39 bp's per month, which on an annual basis is 468 bp's. The posterior mean for β is $p(\beta|D_t) = .745$, so the portfolio takes only 75% of the risk of the market index. In terms of average returns in this period, the market return is 12.25% and the optimized portfolio is 14.05%. This shows significant outperformance of the optimized portfolio over the passive index strategy. In terms of wealth relatives, Figure 2 shows the growth of W_T over time for both portfolios, where $W_T(P) = \prod_{t=1}^{T}(1 + R_i(P))$ and R_i is the return in the ith period for portfolio P.

These results are, of course, sensitive to the choice of portfolio constraints and the length of data history used to form statistical estimates. Black (1976) pointed out the advantage of using daily data for forming volatility predictions. Therefore, one of the valuable uses of high-frequency daily data is that the researcher can rely on a smaller historical data length for forming volatility predictions.
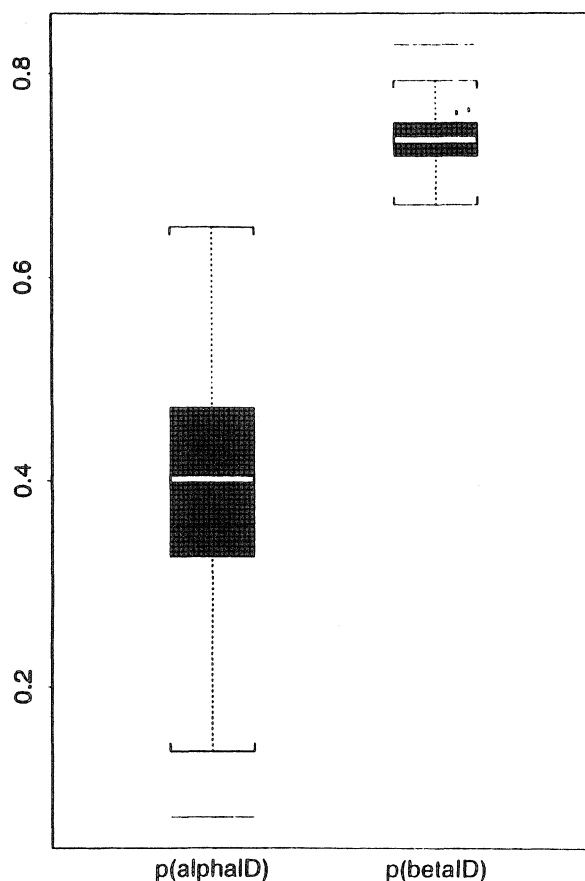
Posterior Distributions

Figure 3. Posterior Distributions.

5. CONCLUSIONS

A predictive approach to portfolio selection has been outlined and implemented. The goal is portfolio selection from a large universe of possible securities. We describe several hierarchical models that can be implemented on high-frequency daily data. The modeling procedure leads to predictive expected returns and a variance–covariance matrix, which is then passed on to an optimization procedure. The predictive minimum-variance portfolio performs well in a back-test against its underlying S&P 500 index benchmark from the period January 1, 1970, to December 31, 1996. There are clearly many areas for future research. The following five issues appear to be of interest:

1. Understanding the interaction between portfolio constraints and the choice of prior $p(\mu, \Sigma)$ in relation to the sensitivity of the optimized weights $x_{MV}(\Sigma_P)$
2. How to calibrate the optimal window for estimation with the holding period for the portfolio
3. Developing models that allow for alternative nonstationarity in returns
4. Understanding what features of a model and historical returns imply zero weights on securities in the optimization
5. Understanding how rebalancing and updating parameter distributions affect the optimal portfolio

In summary, hierarchical modeling and a predictive minimum-variance approach to portfolio selection performs well in large-scale portfolio optimization problems in which performance is measured relative to the benchmark S&P 500 index.

ACKNOWLEDGMENTS

APPENDIX: PROOF OF PROPOSITION 4

Under the model given by $L(\Sigma|\mu, D_t)$, neglecting irrelevant constants,

$$L(\Sigma|\mu, D_t) = (\det(\Sigma_{11}))^{-(s-1)/2}$$

$$\times \exp\left(-\frac{1}{2}\sum_{t=1}^{s-1}(R_{1,t} - \mu)'\Sigma_{11}^{-1}(R_{1,t} - \mu)\right)$$

$$(\det(\Sigma))^{-(T-s)/2}\exp\left(-\frac{1}{2}\sum_{t=s}^{T-s}(R_t - \mu)'\Sigma^{-1}(R_t - \mu)\right).$$

Let $R = 1/(T-s)\sum_{i=s}^{T}(R_t - \mu)(R_t - \mu)'$, so

$$L(\Sigma|\mu, D_t) = (\det(\Sigma_{11}))^{-(s-1)/2}$$

$$\times \exp\left(-\frac{1}{2}\sum_{t=1}^{s-1}(R_{1,t} - \mu)'\Sigma_{11}^{-1}(R_{1,t} - \mu)\right)$$

$$(\det(\Sigma))^{-(T-s)/2}\exp\left(-\frac{1}{2}\operatorname{tr}((T-s)\Sigma^{-1}R)\right).$$

Now consider the decomposition of Σ into its partitioned elements. Under the reparameterization $\Sigma \to (\Sigma_{11}, \psi, \Sigma_{22.1})$, we can compute the likelihood function $L(\Sigma_{11}, \psi, \Sigma_{22.1})$ as follows: Note that

$$\det(\Sigma) = \det(\Sigma_{11})\det(\Sigma_{22.1})$$

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \psi\Sigma_{22.1}^{-1}\psi' & -\Sigma_{22.1}^{-1}\psi \\ -\psi\Sigma_{22.1}^{-1} & \Sigma_{22.1}^{-1} \end{pmatrix}.$$

Let the latter matrix involving ψ and $\Sigma_{22.1}$ be denoted by $A(\psi, \Sigma_{22.1})$. We can rewrite

$$\operatorname{tr}((T-s)\Sigma^{-1}R)$$

$$= \operatorname{tr}((T-s)\Sigma_{11}^{-1}R_{11}) + \operatorname{tr}((T-s)A(\psi, \Sigma_{22.1})R),$$

where we have partitioned R in a similar manner to Σ. Algebraic manipulation and completing the square leads to a factorization of $\operatorname{tr}((T-s)A(\psi, \Sigma_{22.1})R)$ as

$$\operatorname{tr}((T-s)A(\psi, \Sigma_{22.1})R) = \operatorname{tr}((T-s)\Sigma_{22.1}^{-1}R^*(\psi)),$$

where

$$R^*(\psi) = (R_{22.1} + (\psi - R_{11}^{-1}R_{12})'R_{11}(\psi - R_{11}^{-1}R_{12})).$$

The likelihood function $L(\Sigma_{11}, \psi, \Sigma_{22.1}|R)$ is then given by

$$L(\Sigma_{11}, \psi, \Sigma_{22.1}|R) = \det(\Sigma_{11})^{-T/2}\exp\left(-\frac{1}{2}\operatorname{tr}(T\Sigma_{11}^{-1}R_{11}^T)\right)$$

$$\det(\Sigma_{22.1})^{-(T-s)/2}\exp\left(-\frac{1}{2}\operatorname{tr}((T-s)\Sigma_{22.1}^{-1}R^*(\psi))\right). \quad (10)$$

where $R_{11}^T = (1/T)\Sigma_{t=1}^{T}(R_t - \mu)(R_t - \mu)'$. By Bayes theorem, the posterior distribution for Σ is given by

$$p(\Sigma|\mu, D_t) = p(\Sigma_{11}, \psi, \Sigma_{22.1}|\mu, D_t)$$

$$\propto L(\Sigma_{11}, \psi, \Sigma_{22.1}|\mu, D_t)p(\Sigma_{11})p(\psi|\Sigma_{22.1})p(\Sigma_{22.1}).$$

Therefore, the conditional posterior distributions $p(\Sigma_{11}|\mu, D_t), p(\psi|\Sigma_{22.1}, \mu, D_t)$, and $p(\Sigma_{22.1}|\psi, \mu, D_t)$ are all available in closed form and are given by

$$\Sigma_{11}|\mu, D_t \sim W^{-1}\left(\nu_{11} + T, \frac{\nu_{11}}{\nu_{11} + T}\Omega_{11} + \frac{T}{\nu_{11} + T}R_{11}^T\right)$$

$$\psi|\Sigma_{22.1}, \mu, D_t \sim N(\psi_{(T-s)}, \Omega_{(T-s)}, \Sigma_{22.1})$$

$$\Sigma_{22.1}|\psi, \mu, D_t \sim W^{-1}\left(\nu_{22} + (T-s), \right.$$

$$\left. \frac{\nu_{22}}{\nu_{22} + (T-s)}\Omega_{22.1} + \frac{T-s}{\nu_{22} + (T-s)}(R_{22.1} + R^*(\psi))\right)$$

as required.

REFERENCES

Black, F. (1976), "Studies of Stock Market Volatility Changes," in *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pp. 177–181.

Black, F., and Litterman, R. (1993), "Asset Allocation: Combining Investor Views With Market Equilibrium," Research Report, Goldman Sachs, New York.

Brown, S. J. (1979), "The Effect of Estimation Risk on Capital Market Equilibrium," *Journal of Financial and Quantitative Analysis*, 14, 215–220.

Chopra, V. K., and Ziemba, W. T. (1993), "The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice," *Journal of Portfolio Management*, 19, 6–11.

Dawid, A. P. (1981), "Some Matrix-Variate Distribution Theory: Notational Considerations and a Bayesian Approach," *Biometrika*, 68, 265–274.

Dumas, B., and Jacquillant, B. (1990), "Performances of Currency Portfolios Chosen by a Bayesian Technique: 1967–1985," *Journal of Banking and Finance*, 14, 539–558.

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

——— (1988), "For Better Performance: Constrain Portfolio Weights," *Journal of Portfolio Management*, 14, 29–34.

Harris, L., and Gurel, E. (1986), "Price and Volume Effects Associated With Changes in the S&P 500 List: New Evidence for the Existence of Price Pressure," *Journal of Finance*, 41, 815–829.

Jacquier, E., Polson, N., and Rossi, P. (1994), "Bayesian Analysis of Stochastic Volatility Models" (with discussion), *Journal of Business & Economic Statistics*, 12, 371–418.

——— (1995), "Models and Priors for Multivariate Stochastic Volatility Models," technical report, University of Chicago, Graduate School of Business.

Jobson, J. D., and Korkie, B. (1980), "Estimation for Markowitz Efficient Portfolios," *Journal of the American Statistical Association*, 75, 544–554.

Jorion, P. (1986), "Bayes–Stein Estimation for Portfolio Analysis," *Journal of Financial and Quantitative Analysis*, 21, 279–229.

Kandel, S., Stambaugh, R. F., and McCulloch, R. (1996), "Bayesian Inference and Portfolio Efficiency," technical report, National Bureau of Economic Research, Cambridge, MA.

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

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

Michaud, R. O. (1989), "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?" *Financial Analysts Journal*, January, 31–42.

Perold, A. F. (1988), "Large-Scale Portfolio Optimization," *Management Science*, 30, 1143–1160.

Putnam, B. H., and Quintana, J. M. (1994), "New Bayesian Statistical Approaches to Estimating and Evaluating Models of Exchange Rates Determination," in *Proceedings of the Section on Bayesian Statistical Science, American Statistical Association*, pp. 232–237.

Shleifer, A. (1986), "Do Demand Curves for Stocks Slope Downward?" *Journal of Finance*, 41, 579–590.

Stambaugh, R. (1997), "Analyzing Investments Whose Histories Differ in Length," technical report, University of Pennsylvania, Wharton School of Business.

Vasicek, O. (1973), "A Note on Using Cross-sectional Information in Bayesian Estimation of Security Betas," *Journal of Finance*, 28, 1233–1239.

West, M., and Harrison, J. (1989), *Bayesian Forecasting and Dynamic Models*, New York: Springer-Verlag.

Zellner, A., and Chetty, V. (1965), "Prediction and Decision Problems in Regression Models From the Bayesian Point of View," *Journal of the American Statistical Association*, 605–616.