MEAN–VARIANCE PORTFOLIO OPTIMIZATION WHEN MEANS AND COVARIANCES ARE UNKNOWN[1]

By Tze Leung Lai, Haipeng Xing and Zehao Chen

Stanford University, SUNY at Stony Brook and Bosera Fund

Markowitz's celebrated mean–variance portfolio optimization theory assumes that the means and covariances of the underlying asset returns are known. In practice, they are unknown and have to be estimated from historical data. Plugging the estimates into the efficient frontier that assumes known parameters has led to portfolios that may perform poorly and have counter-intuitive asset allocation weights; this has been referred to as the "Markowitz optimization enigma." After reviewing different approaches in the literature to address these difficulties, we explain the root cause of the enigma and propose a new approach to resolve it. Not only is the new approach shown to provide substantial improvements over previous methods, but it also allows flexible modeling to incorporate dynamic features and fundamental analysis of the training sample of historical data, as illustrated in simulation and empirical studies.

1. Introduction. The mean–variance (MV) portfolio optimization theory of Harry Markowitz (1952, 1959), Nobel laureate in economics, is widely regarded as one of the foundational theories in financial economics. It is a single-period theory on the choice of portfolio weights that provide the optimal tradeoff between the mean (as a measure of profit) and the variance (as a measure of risk) of the portfolio return for a future period. The theory, which will be briefly reviewed in the next paragraph, assumes that the means and covariances of the underlying asset returns are known. How to implement the theory in practice when the means and covariances are unknown parameters has been an intriguing statistical problem in financial economics. This paper proposes a novel approach to resolve the long-standing problem and illustrates it with an empirical study using CRSP (the Center for Research in Security Prices of the University of Chicago) monthly stock price

data, which can be accessed via the Wharton Research Data Services at the University of Pennsylvania.

For a portfolio consisting of m assets (e.g., stocks) with expected returns μ_i, let w_i be the weight of the portfolio's value invested in asset i such that $\sum_{i=1}^{m} w_i = 1$, and let $\mathbf{w} = (w_1, \ldots, w_m)^T$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)^T$, $\mathbf{1} = (1, \ldots, 1)^T$. The portfolio return has mean $\mathbf{w}^T \boldsymbol{\mu}$ and variance $\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}$, where $\boldsymbol{\Sigma}$ is the covariance matrix of the asset returns; see Lai and Xing (2008), pages 67, 69–71. Given a target value μ_* for the mean return of a portfolio, Markowitz characterizes an efficient portfolio by its weight vector $\mathbf{w}_{\text{eff}}$ that solves the optimization problem

$$(1.1) \quad \mathbf{w}_{\text{eff}} = \arg \min_{\mathbf{w}} \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^T \boldsymbol{\mu} = \mu_*, \mathbf{w}^T \mathbf{1} = 1, \mathbf{w} \geq \mathbf{0}.$$

When short selling is allowed, the constraint $\mathbf{w} \geq \mathbf{0}$ (i.e., $w_i \geq 0$ for all i) in (1.1) can be removed, yielding the following problem that has an explicit solution:

$$
\begin{aligned}
\mathbf{w}_{\text{eff}} &= \arg \min_{\mathbf{w}: \mathbf{w}^T \boldsymbol{\mu} = \mu_*, \ \mathbf{w}^T \mathbf{1} = 1} \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \\
&= \{ B \boldsymbol{\Sigma}^{-1} \mathbf{1} - A \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \mu_* (C \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - A \boldsymbol{\Sigma}^{-1} \mathbf{1}) \} / D,
\end{aligned}
\tag{1.2}
$$

where $A = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}, B = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}, C = \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}$, and $D = BC - A^2$.

Markowitz's theory assumes known $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Since in practice $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown, a commonly used approach is to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ from historical data, under the assumption that returns are i.i.d. A standard model for the price P_{it} of the ith asset at time t in finance theory is geometric Brownian motion $dP_{it}/P_{it} = \theta_i \, dt + \sigma_i \, dB_t^{(i)}$, where $\{B_t^{(i)}, t \geq 0\}$ is standard Brownian motion. The discrete-time analog of this price process has returns $r_{it} = (P_{it} - P_{i,t-1})/P_{i,t-1}$, and log returns $\log(P_{it}/P_{i,t-1}) = \log(1 + r_{it}) \approx r_{it}$ that are i.i.d. $N(\theta_i - \sigma_i^2/2, \sigma_i^2)$. Under the standard model, maximum likelihood estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the sample mean $\widehat{\boldsymbol{\mu}}$ and the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}$, which are also method-of-moments estimates without the assumption of normality and when the i.i.d. assumption is replaced by weak stationarity (i.e., time-invariant means and covariances). It has been found, however, that replacing $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in (1.1) or (1.2) by their sample counterparts $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ may perform poorly and a major direction in the literature is to find other (e.g., Bayes and shrinkage) estimators that yield better portfolios when they are plugged into (1.1) or (1.2). An alternative method, introduced by Michaud (1989) to tackle the "Markowitz optimization enigma," is to adjust the plug-in portfolio weights by incorporating sampling variability of $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})$ via the bootstrap. Section 2 gives a brief survey of these approaches.

Let $\mathbf{r}_t = (r_{1t}, \ldots, r_{mt})^T$. Since Markowitz's theory deals with portfolio returns in a future period, it is more appropriate to use the conditional mean and covariance matrix of the future returns $\mathbf{r}_{n+1}$ given the historical data $\mathbf{r}_n$, $\mathbf{r}_{n-1}, \ldots$ based on a Bayesian model that forecasts the future from the available data, rather than restricting to an i.i.d. model that relates the future to the past via the unknown parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ for future returns to be estimated from past data. More importantly, this Bayesian formulation paves the way for a new approach that generalizes Markowitz's portfolio theory to the case where the means and covariances are unknown. When $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are estimated from data, their uncertainties should be incorporated into the risk; moreover, it is not possible to attain a target level of mean return as in Markowitz's constraint $\mathbf{w}^T\boldsymbol{\mu} = \mu_*$ since $\boldsymbol{\mu}$ is unknown. To address this root cause of the Markowitz enigma, we introduce in Section 3 a Bayesian approach that assumes a prior distribution for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and formulates mean–variance portfolio optimization as a stochastic optimization problem. This optimization problem reduces to that of Markowitz when the prior distribution is degenerate. It uses the posterior distribution given current and past observations to incorporate the uncertainties of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ into the variance of the portfolio return $\mathbf{w}^T\mathbf{r}_{n+1}$, where $\mathbf{w}$ is based on the posterior distribution. The constraint in Markowitz's mean–variance formulation can be included in the objective function by using a Lagrange multiplier λ^{-1} so that the optimization problem is to evaluate the weight vector $\mathbf{w}$ that maximizes $E(\mathbf{w}^T\mathbf{r}_{n+1}) - \lambda \operatorname{Var}(\mathbf{w}^T\mathbf{r}_{n+1})$, for which λ can be regarded as a risk aversion coefficient. To compare with previous frequentist approaches that assume i.i.d. returns, Section 4 introduces a variant of the Bayes rule that uses bootstrap resampling to estimate the performance criterion nonparametrically.

To apply this theory in practice, the investor has to figure out his/her risk aversion coefficient, which may be a difficult task. Markowitz's theory circumvents this by considering the *efficient frontier*, which is the (σ, μ) curve of efficient portfolios as λ varies over all possible values, where μ is the mean and σ^2 the variance of the portfolio return. Investors, however, often prefer to use $(\mu - \mu_0)/\sigma_e$, called the *information ratio*, as a measure of a portfolio's performance, where μ_0 is the expected return of a benchmark investment and σ_e^2 is the variance of the portfolio's excess return over the benchmark portfolio; see Grinold and Kahn (2000), page 5. The benchmark investment can be a market portfolio (e.g., S&P500) or some other reference portfolio, or a risk-free bank account with interest rate μ_0 (in which case the information ratio is often called the *Sharpe ratio*). Note that the information ratio is proportional to $\mu - \mu_0$ and inversely proportional to σ_e, and can be regarded as the excess return per unit of risk. In Section 5 we describe how λ can be chosen for the rule developed in Section 3 to maximize the information ratio. Other statistical issues that arise in practice are also considered in

Sections 5 and 6 where they lead to certain modifications of the basic rule. Among them are dimension reduction when m (number of assets) is not small relative to n (number of past periods in the training sample) and departures of the historical data from the working assumption of i.i.d. asset returns. Section 6 illustrates these methods in an empirical study in which the rule thus obtained is compared with other rules proposed in the literature. Some concluding remarks are given in Section 7.

2. Using better estimates of μ, Σ or w_{eff} to implement Markowitz's portfolio optimization theory. Since μ and Σ in Markowitz's efficient frontier are actually unknown, a natural idea is to replace them by the sample mean vector $\widehat{\mu}$ and covariance matrix $\widehat{\Sigma}$ of the training sample. However, this plug-in frontier is no longer optimal because $\widehat{\mu}$ and $\widehat{\Sigma}$ actually differ from μ and Σ, and Frankfurter, Phillips and Seagle (1976) and Jobson and Korkie (1980) have reported that portfolios associated with the plug-in frontier can perform worse than an equally weighted portfolio that is highly inefficient. Michaud (1989) comments that the minimum variance (MV) portfolio $\mathbf{w}_{\text{eff}}$ based on $\widehat{\mu}$ and $\widehat{\Sigma}$ has serious deficiencies, calling the MV optimizers "estimation-error maximizers." His argument is reinforced by subsequent studies, for example, Best and Grauer (1991), Chopra, Hensel and Turner (1993), Canner et al. (1997), Simann (1997) and Britten-Jones (1999). Three approaches have emerged to address the difficulty during the past two decades. The first approach uses multifactor models to reduce the dimension in estimating Σ, and the second approach uses Bayes or other shrinkage estimates of Σ. Both approaches use improved estimates of Σ for the plug-in efficient frontier. They have also been modified to provide better estimates of μ, for example, in the quasi-Bayesian approach of Black and Litterman (1990). The third approach uses bootstrapping to correct for the bias of $\widehat{\mathbf{w}}_{\text{eff}}$ as an estimate of $\mathbf{w}_{\text{eff}}$.

2.1. *Multifactor pricing models.* Multifactor pricing models relate the m asset returns r_i to k factors $f_1, \ldots, f_k$ in a regression model of the form

$$(2.1) \qquad r_i = \alpha_i + (f_1, \ldots, f_k)^T \boldsymbol{\beta}_i + \epsilon_i,$$

in which α_i and $\boldsymbol{\beta}_i$ are unknown regression parameters and ϵ_i is an unobserved random disturbance that has mean 0 and is uncorrelated with $\mathbf{f} := (f_1, \ldots, f_k)^T$. The case $k = 1$ is called a *single-factor* (or *single-index*) model. Under Sharpe's (1964) capital asset pricing model (CAPM) which assumes, besides known μ and Σ, that the market has a risk-free asset with return r_f (interest rate) and that all investors minimize the variance of their portfolios for their target mean returns, (2.1) holds with $k = 1$, $\alpha_i = r_f$ and $f = r_M - r_f$, where r_M is the return of a hypothetical market portfolio M

which can be approximated in practice by an index fund such as Standard and Poor's (S&P) 500 Index. The arbitrage pricing theory (APT), introduced by Ross (1976), involves neither a market portfolio nor a risk-free asset and states that a multifactor model of the form (2.1) should hold approximately in the absence of arbitrage for sufficiently large m. The theory, however, does not specify the factors and their number. Methods for choosing factors in (2.1) can be broadly classified as economic and statistical, and commonly used statistical methods include factor analysis and principal component analysis; see Section 3.4 of Lai and Xing (2008).

2.2. *Bayes and shrinkage estimators.* A popular conjugate family of prior distributions for estimation of covariance matrices from i.i.d. normal random vectors $\mathbf{r}_t$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is

$$(2.2) \qquad \boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N(\boldsymbol{\nu}, \boldsymbol{\Sigma}/\kappa), \qquad \boldsymbol{\Sigma} \sim IW_m(\boldsymbol{\Psi}, n_0),$$

where $IW_m(\boldsymbol{\Psi}, n_0)$ denotes the inverted Wishart distribution with n_0 degrees of freedom and mean $\boldsymbol{\Psi}/(n_0 - m - 1)$. The posterior distribution of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ given $(\mathbf{r}_1, \ldots, \mathbf{r}_n)$ is also of the same form:

$$\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N(\widehat{\boldsymbol{\mu}}, \boldsymbol{\Sigma}/(n + \kappa)), \qquad \boldsymbol{\Sigma} \sim IW_m((n + n_0 - m - 1)\widehat{\boldsymbol{\Sigma}}, n + n_0),$$

where $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ are the Bayes estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ given by

$$\widehat{\boldsymbol{\mu}} = \frac{\kappa}{n + \kappa}\boldsymbol{\nu} + \frac{n}{n + \kappa}\bar{\mathbf{r}},$$

$$(2.3) \qquad \widehat{\boldsymbol{\Sigma}} = \frac{n_0 - m - 1}{n + n_0 - m - 1}\frac{\boldsymbol{\Psi}}{n_0 - m - 1}$$

$$+ \frac{n}{n + n_0 - m - 1}\left\{\frac{1}{n}\sum_{i=1}^{n}(\mathbf{r}_t - \bar{\mathbf{r}})(\mathbf{r}_t - \bar{\mathbf{r}})^T\right.$$

$$\left. + \frac{\kappa}{n + \kappa}(\bar{\mathbf{r}} - \boldsymbol{\nu})(\bar{\mathbf{r}} - \boldsymbol{\nu})^T\right\}.$$

Note that the Bayes estimator $\widehat{\boldsymbol{\Sigma}}$ adds to the MLE of $\boldsymbol{\Sigma}$ the covariance matrix $\kappa(\bar{\mathbf{r}} - \boldsymbol{\nu})(\bar{\mathbf{r}} - \boldsymbol{\nu})^T/(n + \kappa)$, which accounts for the uncertainties due to replacing $\boldsymbol{\mu}$ by $\bar{\mathbf{r}}$, besides shrinking this adjusted covariance matrix toward the prior mean $\boldsymbol{\Psi}/(n_0 - m - 1)$.

Simply using $\bar{\mathbf{r}}$ to estimate $\boldsymbol{\mu}$, Ledoit and Wolf (2003, 2004) propose to shrink the MLE of $\boldsymbol{\Sigma}$ toward a structured covariance matrix, instead of using directly this Bayes estimator which requires specification of the hyperparameters $\boldsymbol{\mu}$, κ, n_0 and $\boldsymbol{\Psi}$. Their rationale is that whereas the MLE $\mathbf{S} = \sum_{t=1}^{n}(\mathbf{r}_t - \bar{\mathbf{r}})(\mathbf{r}_t - \bar{\mathbf{r}})^T/n$ has a large estimation error when $m(m + 1)/2$ is comparable with n, a structured covariance matrix $\mathbf{F}$ has much fewer

parameters that can be estimated with smaller variances. They propose to estimate $\boldsymbol{\Sigma}$ by a convex combination of $\widehat{\mathbf{F}}$ and $\mathbf{S}$:

$$(2.4) \qquad \widehat{\boldsymbol{\Sigma}} = \hat{\delta}\widehat{\mathbf{F}} + (1 - \hat{\delta})\mathbf{S},$$

where $\hat{\delta}$ is an estimator of the *optimal shrinkage constant* δ used to shrink the MLE toward the estimated structured covariance matrix $\widehat{\mathbf{F}}$. Besides the covariance matrix $\mathbf{F}$ associated with a single-factor model, they also suggest using a constant correlation model for $\mathbf{F}$ in which all pairwise correlations are identical, and have found that it gives comparable performance in simulation and empirical studies. They advocate using this shrinkage estimate in lieu of $\mathbf{S}$ in implementing Markowitz's efficient frontier.

The difficulty of estimating $\boldsymbol{\mu}$ well enough for the plug-in portfolio to have reliable performance was pointed out by Black and Litterman (1990), who proposed the following pragmatic quasi-Bayesian approach to address this difficulty. Whereas Jorion (1986) had used earlier a shrinkage estimator similar to $\widehat{\boldsymbol{\mu}}$ in (2.3), which can be viewed as shrinking a prior mean $\boldsymbol{\nu}$ to the sample mean $\bar{\mathbf{r}}$ (instead of the other way around), Black and Litterman's approach basically amounted to shrinking an investor's subjective estimate of $\boldsymbol{\mu}$ to the market's estimate implied by an "equilibrium portfolio." The investor's subjective guess of $\boldsymbol{\mu}$ is described in terms of "views" on linear combinations of asset returns, which can be based on past observations and the investor's personal/expert opinions. These views are represented by $\mathbf{P}\boldsymbol{\mu} \sim N(\mathbf{q}, \boldsymbol{\Omega})$, where $\mathbf{P}$ is a $p \times m$ matrix of the investor's "picks" of the assets to express the guesses, and $\boldsymbol{\Omega}$ is a diagonal matrix that expresses the investor's uncertainties in the views via their variances. The "equilibrium portfolio," denoted by $\widetilde{\mathbf{w}}$, is based on a normative theory of an equilibrium market, in which $\widetilde{\mathbf{w}}$ is assumed to solve the mean–variance optimization problem $\max_{\mathbf{w}}(\mathbf{w}^T\boldsymbol{\pi} - \lambda\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w})$, with λ being the average risk-aversion level of the market and $\boldsymbol{\pi}$ representing the market's view of $\boldsymbol{\mu}$. This theory yields the relation $\boldsymbol{\pi} = 2\lambda\boldsymbol{\Sigma}\widetilde{\mathbf{w}}$, which can be used to infer $\boldsymbol{\pi}$ from the market capitalization or benchmark portfolio as a surrogate of $\widetilde{\mathbf{w}}$. Incorporating uncertainty in the market's view of $\boldsymbol{\mu}$, Black and Litterman assume that $\boldsymbol{\pi} - \boldsymbol{\mu} \sim N(\mathbf{0}, \tau\boldsymbol{\Sigma})$, in which $\tau \in (0, 1)$ is a small parameter, and also set exogenously $\lambda = 1.2$; see Meucci (2010). Combining $\mathbf{P}\boldsymbol{\mu} \sim N(\mathbf{q}, \boldsymbol{\Omega})$ with $\boldsymbol{\pi} - \boldsymbol{\mu} \sim N(\mathbf{0}, \tau\boldsymbol{\Sigma})$ under a working independence assumption between the two multivariate normal distributions yields the Black–Litterman estimate of $\boldsymbol{\mu}$:

$$(2.5) \qquad \widehat{\boldsymbol{\mu}}^{\mathrm{BL}} = [(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^T\boldsymbol{\Omega}^{-1}\mathbf{P}]^{-1}[(\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\pi} + \mathbf{P}^T\boldsymbol{\Omega}^{-1}\mathbf{q}],$$

with covariance matrix $[(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^T\boldsymbol{\Omega}^{-1}\mathbf{P}]^{-1}$. Various modifications and extensions of their idea have been proposed; see Meucci (2005), pages 426–437, Fabozzi et al. (2007), pages 232–253, and Meucci (2010). These extensions have the basic form (2.5) or some variant thereof, and differ mainly

in the normative model used to generate an equilibrium portfolio. Note that (2.5) involves $\boldsymbol{\Sigma}$, which Black and Litterman estimated by using the sample covariance matrix of historical data, and that their focus was to address the estimation of $\boldsymbol{\mu}$ for the plug-in portfolio. Clearly Bayes or shrinkage estimates of $\boldsymbol{\Sigma}$ can be used instead.

2.3. *Bootstrapping and the resampled frontier.* To adjust for the bias of $\widehat{\mathbf{w}}_{\text{eff}}$ as an estimate of $\mathbf{w}_{\text{eff}}$, Michaud (1989) uses the average of the bootstrap weight vectors:

$$(2.6) \qquad \bar{\mathbf{w}} = B^{-1} \sum_{b=1}^{B} \widehat{\mathbf{w}}_b^*,$$

where $\widehat{\mathbf{w}}_b^*$ is the estimated optimal portfolio weight vector based on the bth bootstrap sample $\{\mathbf{r}_{b1}^*, \ldots, \mathbf{r}_{bn}^*\}$ drawn with replacement from the observed sample $\{\mathbf{r}_1, \ldots, \mathbf{r}_n\}$. Specifically, the bth bootstrap sample has sample mean vector $\widehat{\boldsymbol{\mu}}_b^*$ and covariance matrix $\widehat{\boldsymbol{\Sigma}}_b^*$, which can be used to replace $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in (1.1) or (1.2), thereby yielding $\widehat{\mathbf{w}}_b^*$. Thus, the resampled efficient frontier corresponds to plotting $\bar{\mathbf{w}}^T \widehat{\boldsymbol{\mu}}$ versus $\sqrt{\bar{\mathbf{w}}^T \widehat{\boldsymbol{\Sigma}} \bar{\mathbf{w}}}$ for a fine grid of μ_* values, where $\bar{\mathbf{w}}$ is defined by (2.6) in which $\widehat{\mathbf{w}}_b^*$ depends on the target level $\boldsymbol{\mu}_*$.

3. A stochastic optimization approach. The Bayesian and shrinkage methods in Section 2.2 focus primarily on Bayes estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ (with normal and inverted Wishart priors) and shrinkage estimators of $\boldsymbol{\Sigma}$. However, the construction of efficient portfolios when $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown is more complicated than trying to estimate them as well as possible and then plugging the estimates into (1.1) or (1.2). Note in this connection that (1.2) involves $\boldsymbol{\Sigma}^{-1}$ instead of $\boldsymbol{\Sigma}$ and that estimating $\boldsymbol{\Sigma}$ as well as possible does not imply that $\boldsymbol{\Sigma}^{-1}$ is reliably estimated. Estimation of a high-dimensional $m \times m$ covariance matrix and its inverse when m^2 is not small compared to n has been recognized as a difficult statistical problem and attracted much recent attention; see, for example, Ledoit and Wolf (2004), Huang et al. (2006), Bickel and Lavina (2008) and Fan, Fan and Lv (2008). Some sparsity condition or a low-dimensional factor structure is needed to obtain an estimate which is close to $\boldsymbol{\Sigma}$ and whose inverse is close to $\boldsymbol{\Sigma}^{-1}$, but the conjugate prior family (2.2) that motivates the (linear) shrinkage estimators (2.3) or (2.4) does not reflect such sparsity. For high-dimensional weight vectors $\widehat{\mathbf{w}}_{\text{eff}}$, direct application of the bootstrap for bias correction is also problematic.

A major difficulty with the "plug-in" efficient frontier (which uses $\mathbf{S}$ to estimate $\boldsymbol{\Sigma}$ and $\bar{\mathbf{r}}$ to estimate $\boldsymbol{\mu}$), its variants that estimate $\boldsymbol{\Sigma}$ by (2.4) and $\boldsymbol{\mu}$ by (2.3) or the Black–Litterman method, and its "resampled" version is that Markowitz's idea of using the variance of $\mathbf{w}^T \mathbf{r}_{n+1}$ as a measure of the

portfolio's risk cannot be captured simply by the plug-in estimates $\mathbf{w}^T\widehat{\boldsymbol{\Sigma}}\mathbf{w}$ of $\mathrm{Var}(\mathbf{w}^T\mathbf{r}_{n+1})$ and $\mathbf{w}^T\widehat{\boldsymbol{\mu}}$ of $E(\mathbf{w}^T\mathbf{r}_{n+1})$. This difficulty was recognized by Broadie (1993), who used the terms *true frontier* and *estimated frontier* to refer to Markowitz's efficient frontier (with known $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$) and the plug-in efficient frontier, respectively, and who also suggested considering the *actual* mean and variance of the return of an estimated frontier portfolio. Whereas the problem of minimizing $\mathrm{Var}(\mathbf{w}^T\mathbf{r}_{n+1})$ subject to a given level μ_* of the mean return $E(\mathbf{w}^T\mathbf{r}_{n+1})$ is meaningful in Markowitz's framework, in which both $E(\mathbf{r}_{n+1})$ and $\mathrm{Cov}(\mathbf{r}_{n+1})$ are known, the surrogate problem of minimizing $\mathbf{w}^T\widehat{\boldsymbol{\Sigma}}\mathbf{w}$ under the constraint $\mathbf{w}^T\widehat{\boldsymbol{\mu}} = \mu_*$ ignores the fact that both $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ have inherent errors (risks) themselves. In this section we consider the more fundamental problem

$$(3.1) \qquad \max\{E(\mathbf{w}^T\mathbf{r}_{n+1}) - \lambda\,\mathrm{Var}(\mathbf{w}^T\mathbf{r}_{n+1})\}$$

when $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown and treated as state variables whose uncertainties are specified by their posterior distributions given the observations $\mathbf{r}_1,\ldots,\mathbf{r}_n$ in a Bayesian framework. The weights $\mathbf{w}$ in (3.1) are random vectors that depend on $\mathbf{r}_1,\ldots,\mathbf{r}_n$. Note that if the prior distribution puts all its mass at $(\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)$, then the minimization problem (3.1) reduces to Markowitz's portfolio optimization problem that assumes $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are given. The Lagrange multiplier λ in (3.1) can be regarded as the investor's risk-aversion index when variance is used to measure risk.

3.1. *Solution of the optimization problem (3.1).* The problem (3.1) is not a standard stochastic optimization problem because of the term $[E(\mathbf{w}^T\mathbf{r}_{n+1})]^2$ in $\mathrm{Var}(\mathbf{w}^T\mathbf{r}_{n+1}) = E[(\mathbf{w}^T\mathbf{r}_{n+1})^2] - [E(\mathbf{w}^T\mathbf{r}_{n+1})]^2$. A standard stochastic optimization problem in the Bayesian setting is of the form $\max_{a\in\mathcal{A}} Eg(\mathbf{X},\boldsymbol{\theta},a)$, in which $g(\mathbf{X},\boldsymbol{\theta},a)$ is the reward when action a is taken, $\mathbf{X}$ is a random vector with distribution $F_{\boldsymbol{\theta}}$, $\boldsymbol{\theta}$ has a prior distribution and the maximization is over the action space $\mathcal{A}$. The key to its solution is the law of conditional expectations $Eg(\mathbf{X},\boldsymbol{\theta},a) = E\{E[g(\mathbf{X},\boldsymbol{\theta},a)|\mathbf{X}]\}$, which implies that the stochastic optimization problem can be solved by choosing a to maximize the posterior reward $E[g(\mathbf{X},\boldsymbol{\theta},a)|\mathbf{X}]\}$. This key idea, however, cannot be applied to the problem of maximizing or minimizing nonlinear functions of $Eg(\mathbf{X},\boldsymbol{\theta},a)$, such as $[Eg(\mathbf{X},\boldsymbol{\theta},a)]^2$ that is involved in (3.1).

Our method of solving (3.1) is to convert it to a standard stochastic control problem by using an additional parameter. Let $W = \mathbf{w}^T\mathbf{r}_{n+1}$ and note that $E(W) - \lambda\,\mathrm{Var}(W) = h(EW, EW^2)$, where $h(x,y) = x + \lambda x^2 - \lambda y$. Let $W_B = \mathbf{w}_B^T\mathbf{r}_{n+1}$ and $\eta = 1 + 2\lambda E(W_B)$, where $\mathbf{w}_B$ is the Bayes weight vector. Then

$$0 \geq h(EW, EW^2) - h(EW_B, EW_B^2)$$
$$= E(W) - E(W_B) - \lambda\{E(W^2) - E(W_B^2)\} + \lambda\{(EW)^2 - (EW_B)^2\}$$

$$= \eta\{E(W) - E(W_B)\} + \lambda\{E(W_B^2) - E(W^2)\} + \lambda\{E(W) - E(W_B)\}^2$$
$$\geq \{\lambda E(W_B^2) - \eta E(W_B)\} - \{\lambda E(W^2) - \eta E(W)\}.$$

Moreover, the last inequality is strict unless $EW = EW_B$, in which case the first inequality is strict unless $EW^2 = EW_B^2$. This shows that the last term above is ≤ 0 or, equivalently,

$$(3.2) \qquad \lambda E(W^2) - \eta E(W) \geq \lambda E(W_B^2) - \eta E(W_B),$$

and that equality holds in (3.2) if and only if W has the same mean and variance as W_B. Hence, the stochastic optimization problem (3.1) is equivalent to minimizing $\lambda E[(\mathbf{w}^T\mathbf{r}_{n+1})^2] - \eta E(\mathbf{w}^T\mathbf{r}_{n+1})$ over weight vectors $\mathbf{w}$ that can depend on $\mathbf{r}_1,\ldots,\mathbf{r}_n$. Since $\eta = 1 + 2\lambda E(W_B)$ is a linear function of the solution of (3.1), we cannot apply this equivalence directly to the unknown η. Instead we solve a family of standard stochastic optimization problems over η and then choose the η that maximizes the reward in (3.1).

To summarize, we can solve (3.1) by rewriting it as the following maximization problem over η:

$$(3.3) \qquad \max_{\eta}\{E[\mathbf{w}^T(\eta)\mathbf{r}_{n+1}] - \lambda\operatorname{Var}[\mathbf{w}^T(\eta)\mathbf{r}_{n+1}]\},$$

where $\mathbf{w}(\eta)$ is the solution of the stochastic optimization problem

$$\mathbf{w}(\eta) = \arg\min_{\mathbf{w}}\{\lambda E[(\mathbf{w}^T\mathbf{r}_{n+1})^2] - \eta E(\mathbf{w}^T\mathbf{r}_{n+1})\}.$$

3.2. *Computation of the optimal weight vector.* Let $\boldsymbol{\mu}_n$ and $\mathbf{V}_n$ be the posterior mean and second moment matrix given the set $\mathcal{R}_n$ of current and past returns $\mathbf{r}_1,\ldots,\mathbf{r}_n$. Since $\mathbf{w}$ is based on $\mathcal{R}_n$, it follows from $E(\mathbf{r}_{n+1}|\mathcal{R}_n) = \boldsymbol{\mu}_n$ and $E(\mathbf{r}_{n+1}\mathbf{r}_{n+1}^T|\mathcal{R}_n) = \mathbf{V}_n$ that

$$(3.4) \qquad E(\mathbf{w}^T\mathbf{r}_{n+1}) = E(\mathbf{w}^T\boldsymbol{\mu}_n), \qquad E[(\mathbf{w}^T\mathbf{r}_{n+1})^2] = E(\mathbf{w}^T\mathbf{V}_n\mathbf{w}).$$

Without short selling, the weight vector $\mathbf{w}(\eta)$ in (3.3) is given by the following analog of (1.1):

$$(3.5) \qquad \mathbf{w}(\eta) = \arg\min_{\mathbf{w}:\mathbf{w}^T\mathbf{1}=1,\mathbf{w}\geq\mathbf{0}}\{\lambda\mathbf{w}^T\mathbf{V}_n\mathbf{w} - \eta\mathbf{w}^T\boldsymbol{\mu}_n\},$$

which can be computed by quadratic programming (e.g., by `quadprog` in `MATLAB`). When short selling is allowed but there are limits on short setting, the constraint $\mathbf{w} \geq \mathbf{0}$ can be replaced by $\mathbf{w} \geq \mathbf{w}_0$, where $\mathbf{w}_0$ is a vector of negative numbers. When there is no limit on short selling, the constraint $\mathbf{w} \geq \mathbf{0}$ in (3.5) can be removed and $\mathbf{w}(\eta)$ in (3.3) is given explicitly by

$$\mathbf{w}(\eta) = \arg\min_{\mathbf{w}:\mathbf{w}^T\mathbf{1}=1}\{\lambda\mathbf{w}^T\mathbf{V}_n\mathbf{w} - \eta\mathbf{w}^T\boldsymbol{\mu}_n\}$$

$$(3.6)$$

$$= \frac{1}{C_n}\mathbf{V}_n^{-1}\mathbf{1} + \frac{\eta}{2\lambda}\mathbf{V}_n^{-1}\left(\boldsymbol{\mu}_n - \frac{A_n}{C_n}\mathbf{1}\right),$$

where the second equality can be derived by using a Lagrange multiplier and

(3.7) $A_n = \boldsymbol{\mu}_n^T \mathbf{V}_n^{-1} \mathbf{1} = \mathbf{1}^T \mathbf{V}_n^{-1} \boldsymbol{\mu}_n, \qquad B_n = \boldsymbol{\mu}_n^T \mathbf{V}_n^{-1} \boldsymbol{\mu}_n, \qquad C_n = \mathbf{1}^T \mathbf{V}_n^{-1} \mathbf{1}.$

Quadratic programming can be used to compute $\mathbf{w}(\eta)$ for more general linear and quadratic constraints than those in (3.5); see Fabozzi et al. (2007), pages 88–92.

Note that (3.5) or (3.6) essentially plugs the Bayes estimates of $\boldsymbol{\mu}$ and $\mathbf{V} := \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T$ into the optimal weight vector that assumes $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ to be known. However, unlike the "plug-in" efficient frontier described in the first paragraph of Section 2, we have first transformed the original mean–variance portfolio optimization problem into a "mean versus second moment" optimization problem that has an additional parameter η. Putting (3.5) or (3.6) into

(3.8) $C(\eta) := E[\mathbf{w}^T(\eta)\boldsymbol{\mu}_n] + \lambda(E[\mathbf{w}^T(\eta)\boldsymbol{\mu}_n])^2 - \lambda E[\mathbf{w}^T(\eta)\mathbf{V}_n\mathbf{w}(\eta)],$

which is equal to $E[\mathbf{w}^T(\eta)\mathbf{r}] - \lambda \operatorname{Var}[\mathbf{w}^T(\eta)\mathbf{r}]$ by (3.4), we can use Brent's method [Press et al. (1992), pages 359–362] to maximize $C(\eta)$. It should be noted that this argument implicitly assumes that the maximum of (3.1) is attained by some $\mathbf{w}$ and is finite. Whereas this assumption is satisfied when there are limits on short selling as in (3.5), it may not hold when there is no limit on short selling. In fact, the explicit formula of $\mathbf{w}(\eta)$ in (3.6) can be used to express (3.8) as a quadratic function of η:

$$C(\eta) = \frac{\eta^2}{4\lambda} E\left\{\left(B_n - \frac{A_n^2}{C_n}\right)\left(B_n - \frac{A_n^2}{C_n} - 1\right)\right\} + \eta E\left\{\left(\frac{1}{2\lambda} + \frac{A_n}{C_n}\right)\left(B_n - \frac{A_n^2}{C_n}\right)\right\}$$
$$+ E\left\{\frac{A_n}{C_n} + \lambda\frac{A_n^2 - C_n}{C_n^2}\right\},$$

which has a maximum only if

(3.9) $E\left\{\left(B_n - \frac{A_n^2}{C_n}\right)\left(B_n - \frac{A_n^2}{C_n} - 1\right)\right\} < 0.$

In the case $E\{(B_n - \frac{A_n^2}{C_n})(B_n - \frac{A_n^2}{C_n} - 1)\} > 0$, $C(\eta)$ has a minimum instead and approaches to ∞ as $|\eta| \to \infty$. In this case, (3.1) has an infinite value and should be defined as a supremum (which is not attained) instead of a maximum.

REMARK.   Let $\boldsymbol{\Sigma}_n$ denote the posterior covariance matrix given $\mathcal{R}_n$. Note that the law of iterated conditional expectations, from which (3.4) follows, has the following analog for $\operatorname{Var}(W)$:

(3.10)
$$\operatorname{Var}(W) = E[\operatorname{Var}(W|\mathcal{R}_n)] + \operatorname{Var}[E(W|\mathcal{R}_n)]$$
$$= E(\mathbf{w}^T\boldsymbol{\Sigma}_n\mathbf{w}) + \operatorname{Var}(\mathbf{w}^T\boldsymbol{\mu}_n).$$

Using $\boldsymbol{\Sigma}_n$ to replace $\boldsymbol{\Sigma}$ in the optimal weight vector that assumes $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ to be known, therefore, ignores the variance of $\mathbf{w}^T\boldsymbol{\mu}_n$ in (3.10), and this omission is an important root cause for the Markowitz optimization enigma related to "plug-in" efficient frontiers.

4. Empirical Bayes, bootstrap approximation and frequentist risk. For more flexible modeling, one can allow the prior distribution in the preceding Bayesian approach to include unspecified hyperparameters, which can be estimated from the training sample by maximum likelihood, or method of moments or other methods. For example, for the conjugate prior (2.2), we can assume $\boldsymbol{\nu}$ and $\boldsymbol{\Psi}$ to be functions of certain hyperparameters that are associated with a multifactor model of the type (2.1). This amounts to using an empirical Bayes model for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in the stochastic optimization problem (3.1). Besides a prior distribution for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, (3.1) also requires specification of the common distribution of the i.i.d. returns to evaluate $E_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{w}^T\mathbf{r}_{n+1})$ and $\mathrm{Var}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{w}^T\mathbf{r}_{n+1})$. The bootstrap provides a nonparametric method to evaluate these quantities, as described below.

4.1. *Bootstrap estimate of performance.* To begin with, note that we can evaluate the frequentist performance of asset allocation rules by making use of the bootstrap method. The bootstrap samples $\{\mathbf{r}^*_{b1}, \ldots, \mathbf{r}^*_{bn}\}$ drawn with replacement from the observed sample $\{\mathbf{r}_1, \ldots, \mathbf{r}_n\}$, $1 \leq b \leq B$, can be used to estimate its $E_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{w}_n^T\mathbf{r}_{n+1}) = E_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{w}_n^T\boldsymbol{\mu})$ and $\mathrm{Var}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{w}_n^T\mathbf{r}_{n+1}) = E_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{w}_n^T\boldsymbol{\Sigma}\mathbf{w}_n) + \mathrm{Var}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{w}_n^T\boldsymbol{\mu})$ of various portfolios Π whose weight vectors $\mathbf{w}_n$ may depend on $\mathbf{r}_1, \ldots, \mathbf{r}_n$. In particular, we can use Bayes or other estimators for $\boldsymbol{\mu}_n$ and $\mathbf{V}_n$ in (3.5) or (3.6) and then choose η to maximize the bootstrap estimate of $E_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{w}_n^T\mathbf{r}_{n+1}) - \lambda\,\mathrm{Var}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{w}_n^T\mathbf{r}_{n+1})$. This is tantamount to using the empirical distribution of $\mathbf{r}_1, \ldots, \mathbf{r}_n$ to be the common distribution of the returns. In particular, using $\bar{\mathbf{r}}$ for $\boldsymbol{\mu}_n$ in (3.5) and the second moment matrix $n^{-1}\sum_{t=1}^{n}\mathbf{r}_t\mathbf{r}_t^T$ for $\mathbf{V}_n$ in (3.6) provides a "nonparametric empirical Bayes" variant, abbreviated by NPEB hereafter, of the optimal rule in Section 3.

4.2. *A simulation study of Bayes and frequentist rewards.* The following simulation study assumes i.i.d. annual returns (in %) of $m = 4$ assets whose mean vector and covariance matrix are generated from the normal and inverted Wishart prior distribution (2.2) with $\kappa = 5$, $n_0 = 10$, $\boldsymbol{\nu} = (2.48, 2.17, 1.61, 3.42)^T$ and the hyperparameter $\boldsymbol{\Psi}$ given by

$$\boldsymbol{\Psi}_{11} = 3.37, \qquad \boldsymbol{\Psi}_{22} = 4.22, \qquad \boldsymbol{\Psi}_{33} = 2.75, \qquad \boldsymbol{\Psi}_{44} = 8.43,$$

$$\boldsymbol{\Psi}_{12} = 2.04,$$

$$\boldsymbol{\Psi}_{13} = 0.32, \qquad \boldsymbol{\Psi}_{14} = 1.59, \qquad \boldsymbol{\Psi}_{23} = -0.05,$$

$$\boldsymbol{\Psi}_{24} = 3.02, \qquad \boldsymbol{\Psi}_{34} = 1.08.$$

TABLE 1
Rewards of four portfolios formed from $m = 4$ assets

λ	(μ, Σ)	Bayes	Plug-in	Oracle	NPEB
1	Bayes	0.0324 (2.47e−5)	0.0317 (2.55e−5)	0.0328 (2.27e−5)	0.0324 (2.01e−5)
	Freq 1	0.0332 (2.61e−6)	0.0324 (5.62e−6)	0.0332	0.0332 (2.56e−6)
	Freq 2	0.0293 (7.23e−6)	0.0282 (5.32e−6)	0.0298	0.0293 (7.12e−6)
	Freq 3	0.0267 (4.54e−6)	0.0257 (5.57e−6)	0.0268	0.0267 (4.73e−6)
5	Bayes	0.0262 (2.33e−5)	0.0189 (1.21e−5)	0.0267 (2.02e−5)	0.0262 (1.89e−5)
	Freq 1	0.0272 (4.06e−6)	0.0182 (5.54e−6)	0.0273	0.0272 (2.60e−6)
	Freq 2	0.0233 (9.35e−6)	0.0183 (3.88e−6)	0.0240	0.0234 (1.03e−5)
	Freq 3	0.0235 (5.25e−6)	0.0159 (2.88e−6)	0.0237	0.0235 (5.27e−6)
10	Bayes	0.0184 (2.54e−5)	0.0067 (7.16e−6)	0.0190 (2.08e−5)	0.0183 (2.23e−5)
	Freq 1	0.0197 (7.95e−6)	0.0063 (3.63e−6)	0.0199	0.0198 (4.19e−6)
	Freq 2	0.0157 (1.08e−5)	0.0072 (3.00e−6)	0.0168	0.0159 (1.13e−5)
	Freq 3	0.0195 (6.59e−6)	0.0083 (1.62e−6)	0.0198	0.0196 (5.95e−6)

We consider four scenarios for the case $n = 6$ without short selling. The first scenario assumes this prior distribution and studies the Bayesian reward for $\lambda = 1, 5$ and 10. The other scenarios consider the frequentist reward at three values of (μ, Σ) generated from the prior distribution. These values, denoted by Freq 1, Freq 2, Freq 3, are as follows:

Freq 1: $\mu = (2.42, 1.88, 1.58, 3.47)^T$, $\Sigma_{11} = 1.17, \Sigma_{22} = 0.82, \Sigma_{33} = 1.37$, $\Sigma_{44} = 2.86, \Sigma_{12} = 0.79, \Sigma_{13} = 0.84, \Sigma_{14} = 1.61, \Sigma_{23} = 0.61, \Sigma_{24} = 1.23, \Sigma_{34} = 1.35$.

Freq 2: $\mu = (2.59, 2.29, 1.25, 3.13)^T$, $\Sigma_{11} = 1.32, \Sigma_{22} = 0.67, \Sigma_{33} = 1.43$, $\Sigma_{44} = 1.03, \Sigma_{12} = 0.75, \Sigma_{13} = 0.85, \Sigma_{14} = 0.68, \Sigma_{23} = 0.32, \Sigma_{24} = 0.44, \Sigma_{34} = 0.61$.

Freq 3: $\mu = (1.91, 1.58, 1.03, 2.76)^T$, $\Sigma_{11} = 1.00, \Sigma_{22} = 0.83, \Sigma_{33} = 0.35$, $\Sigma_{44} = 0.62, \Sigma_{12} = 0.73, \Sigma_{13} = 0.26, \Sigma_{14} = 0.36, \Sigma_{23} = 0.16, \Sigma_{24} = 0.50, \Sigma_{34} = 0.14$.

Table 1 compares the Bayes rule that maximizes (3.1), called "Bayes" hereafter, with three other rules: (a) the "oracle" rule that assumes μ and Σ to be known, (b) the plug-in rule that replaces μ and Σ by the sample estimates of μ and Σ, and (c) the NPEB (nonparametric empirical Bayes) rule described in Section 4.1 Note that although both (b) and (c) use the sample mean vector and sample covariance (or second moment) matrix, (b) simply plugs the sample estimates into the oracle rule while (c) uses the empirical distribution to replace the common distribution of the returns in the Bayes rule. For the plug-in rule, the quadratic programming procedure may have numerical difficulties if the sample covariance matrix is nearly singular. If it should happen, we use the default option of adding $0.005\mathbf{I}$ to the sample

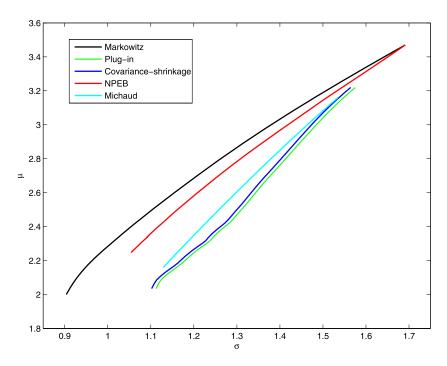FIG. 1. (σ, μ) curves of different portfolios.

covariance matrix. Each result in Table 1 is based on 500 simulations, and the standard errors are given in parentheses. In each scenario, the reward of the NPEB rule is close to that of the Bayes rule and somewhat smaller than that of the oracle rule. The plug-in rule has substantially smaller rewards, especially for larger values of λ.

4.3. *Comparison of the (σ, μ) plots of different portfolios.* The set of points in the (σ, μ) plane that correspond to the returns of portfolios of the m assets is called the *feasible region*. As λ varies over $(0, \infty)$, the (σ, μ) values of the oracle rule correspond to Markowitz's efficient frontier which assumes known $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and which is the upper left boundary of the feasible region. For portfolios whose weights do not assume knowledge of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, the (σ, μ) values lie on the right of Markowitz's efficient frontier. Figure 1 plots the (σ, μ) values of different portfolios formed from $m = 4$ assets without short selling and a training sample of size $n = 6$ when $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by the frequentist scenario Freq 1 above. Markowitz's efficient frontier is computed analytically by varying μ_* in (1.1) over a grid of values. The (σ, μ) curves of the plug-in, covariance-shrinkage [Ledoit and Wolf (2004)] and Michaud's resampled portfolios are computed by Monte Carlo, using 500 simulated paths, for each value of μ_* in a grid ranging from 2.0 to 3.47. The (σ, μ) curve

of the NPEB portfolio is also obtained by Monte Carlo simulations with 500 runs, by using different values of $\lambda > 0$ in a grid. This curve is relatively close to Markowitz's efficient frontier among the (σ, μ) curves of various portfolios that do not assume knowledge of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, as shown in Figure 1. For the covariance-shrinkage portfolio, we use a constant correlation model for $\widehat{F}$ in (2.4), which can be implemented by their software available at www.ledoit.net. Note that Markowitz's efficient frontier has μ values ranging from 2.0 to 3.47, which is the largest component of $\boldsymbol{\mu}$ in Freq 1. The (σ, μ) curve of NPEB lies below the efficient frontier, and further below are the (σ, μ) curves of Michaud's, covariance-shrinkage and plug-in portfolios, in decreasing order. These (σ, μ) curves are what Broadie (1993) calls the *actual frontiers*.

The highest values 3.22, 3.22 and 3.16 of μ for the plug-in, covariance-shrinkage and Michaud's portfolios in Figure 1 are attained with a target value $\mu_* = 3.47$, and the corresponding values of σ are 1.54, 1.54 and 3.16, respectively. Note that without short selling, the constraint $\mathbf{w}^T \widehat{\boldsymbol{\mu}} = \mu_*$ used in these portfolios cannot hold if $\max_{1 \le i \le 4} \widehat{\mu}_i < \mu_*$. We therefore need a default option, such as replacing μ_* by $\min(\mu_*, \max_{1 \le i \le 4} \widehat{\mu}_i)$, to implement the optimization procedures for these portfolios. In contrast, the NPEB portfolio can always be implemented for any given value of λ. In particular, for $\lambda = 0.001$, the NPEB portfolio has $\mu = 3.470$ and $\sigma = 1.691$.

5. Connecting theory to practice. While Section 4 has considered practical implementation of the theory in Section 3, we develop the methodology further in this section to connect the basic theory to practice.

5.1. *The information ratios and choice of* λ. As pointed out in Section 1, the λ in Section 3 is related to how risk-averse one is when one tries to maximize the expected utility of a portfolio. It represents a penalty on the risk that is measured by the variance of the portfolio's return. In practice, it may be difficult to specify an investor's risk aversion parameter λ that is needed in the theory in Section 3.1. A commonly used performance measure of a portfolio's performance is the information ratio $(\mu - \mu_0)/\sigma_e$, which is the excess return per unit of risk; the excess is measured by $\mu - \mu_0$, where $\mu_0 = E(r_0)$, r_0 is the return of the benchmark investment and σ_e^2 is the variance of the excess return. We can regard λ as a tuning parameter, and choose it to maximize the information ratio by modifying the NPEB procedure in Section 3.2, where the bootstrap estimate of $E_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}[\mathbf{w}^T(\eta)\mathbf{r}] - \lambda \operatorname{Var}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}[\mathbf{w}^T(\eta)\mathbf{r}]$ is used to find the portfolio weight $\mathbf{w}_\lambda$ that solves the optimization problem (3.3). Specifically, we use the bootstrap estimate of the information ratio

$$(5.1) \qquad E_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{w}_\lambda \mathbf{r} - r_0)/\sqrt{\operatorname{Var}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{w}_\lambda^T \mathbf{r} - r_0)}$$

of $\mathbf{w}_\lambda$, and maximize the estimated information ratios over λ in a grid that will be illustrated in Section 6.

5.2. *Dimension reduction when m is not small relative to n.* Another statistical issue encountered in practice is the large number m of assets relative to the number n of past periods in the training sample, making it difficult to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ satisfactorily. Using factor models that are related to domain knowledge as in Section 2.1 helps reduce the number of parameters to be estimated in an empirical Bayes approach.

An obvious way of dimension reduction when there is no short selling is to exclude assets with markedly inferior information ratios from consideration. The only potential advantage of including them in the portfolio is that they may be able to reduce the portfolio variance if they are negatively correlated with the "superior" assets. However, since the correlations are unknown, such advantage is unlikely when they are not estimated well enough. Suppose we include in the simulation study of Section 4.2 two more assets so that all asset returns are jointly normal. The additional hyperparameters of the normal and inverted Wishart prior distribution (2.2) are $\boldsymbol{\nu}_5 = -0.014$, $\boldsymbol{\nu}_6 = -0.064$, $\boldsymbol{\Psi}_{55} = 2.02$, $\boldsymbol{\Psi}_{66} = 10.32$, $\boldsymbol{\Psi}_{56} = 0.90$, $\boldsymbol{\Psi}_{15} = -0.17$, $\boldsymbol{\Psi}_{25} = -0.03$, $\boldsymbol{\Psi}_{35} = -0.91$, $\boldsymbol{\Psi}_{45} = -0.33$, $\boldsymbol{\Psi}_{16} = -3.40$, $\boldsymbol{\Psi}_{26} = -3.99$, $\boldsymbol{\Psi}_{36} = -0.08$ and $\boldsymbol{\Psi}_{46} = -3.58$. As in Section 4.2, we consider four scenarios for the case of $n = 8$ without short selling, the first of which assumes this prior distribution and studies the Bayesian reward for $\lambda = 1, 5$ and 10. Table 2 shows the rewards for the four rules in Section 4.2, and each result is based on 500 simulations. Note that the value of the reward function does not show significant change with the inclusion of two additional stocks, which have negative correlations with the four stocks in Section 4.2 but have low information ratios. This shows that excluding stocks with markedly inferior information ratios when there is no short selling can reduce m substantially in practice. In Section 6 we describe another way of choosing stocks from a universe of available stocks to reduce m.

TABLE 2
Rewards of four portfolios formed from $m = 6$ assets

λ	(μ,Σ)	Bayes	Plug-in	Oracle	NPEB
1	Bayes	0.0325 (2.55e−5)	0.0318 (2.62e−6)	0.0331 (2.42e−5)	0.0325 (2.53e−5)
	Freq 1	0.0284 (1.59e−5)	0.0277 (1.31e−5)	0.0296	0.0285 (1.62e−5)
	Freq 2	0.0292 (8.30e−6)	0.0280 (7.95e−6)	0.0296	0.0292 (8.29e−6)
	Freq 3	0.0283 (1.00e−5)	0.0282 (9.11e−6)	0.0300	0.0283 (1.05e−5)
5	Bayes	0.0255 (2.46e−5)	0.0183 (1.44e−5)	0.0263 (2.05e−5)	0.0254 (2.45e−5)
	Freq 1	0.0236 (1.99e−5)	0.0149 (6.48e−6)	0.0250	0.0237 (2.17e−5)
	Freq 2	0.0241 (9.34e−6)	0.0166 (3.61e−6)	0.0246	0.0243 (8.95e−6)
	Freq 3	0.0189 (2.09e−5)	0.0138 (1.45e−5)	0.0219	0.0208 (2.32e−5)
10	Bayes	0.0171 (2.63e−5)	0.0039 (1.57e−5)	0.0180 (2.20e−5)	0.0171 (2.72e−5)
	Freq 1	0.0174 (2.06e−5)	0.0042 (5.19e−6)	0.0193	0.0177 (2.42e−5)
	Freq 2	0.0177 (1.12e−5)	0.0052 (6.34e−6)	0.0184	0.0180 (1.10e−5)
	Freq 3	0.0089 (2.79e−5)	0.0024 (1.33e−5)	0.0120	0.0094 (4.65e−5)

5.3. *Extension to time series models of returns.* An important assumption in the modification of Markowitz's theory in Section 3.2 is that $\mathbf{r}_t$ are i.i.d. with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Diagnostic checks of the extent to which this assumption is violated should be carried out in practice. The stochastic optimization theory in Section 3.1 does not actually need this assumption and only requires the posterior mean and second moment matrix of the return vector for the next period in (3.4). Therefore, one can modify the "working i.i.d. model" accordingly when the diagnostic checks reveal such modifications are needed.

A simple method to introduce such modification is to use a stochastic regression model of the form

$$r_{it} = \boldsymbol{\beta}_i^T \mathbf{x}_{i,t-1} + \epsilon_{it}, \tag{5.2}$$

where the components of $\mathbf{x}_{i,t-1}$ include 1, factor variables such as the return of a market portfolio like S&P500 at time $t-1$, and lagged variables $r_{i,t-1}, r_{i,t-2}, \ldots$. The basic idea underlying (5.2) is to introduce covariates (including lagged variables to account for time series effects) so that the errors ϵ_{it} can be regarded as i.i.d., as in the working i.i.d. model. The regression parameter $\boldsymbol{\beta}_i$ can be estimated by the method of moments, which is equivalent to least squares. We can also include heteroskedasticity by assuming that $\epsilon_{it} = s_{i,t-1}(\boldsymbol{\gamma}_i)z_{it}$, where z_{it} are i.i.d. with mean 0 and variance 1, $\boldsymbol{\gamma}_i$ is a parameter vector which can be estimated by maximum likelihood or generalized method of moments, and $s_{i,t-1}$ is a given function that depends on $r_{i,t-1}, r_{i,t-2}, \ldots$. A well-known example is the GARCH(1,1) model

$$\epsilon_{it} = s_{i,t-1}z_{it}, \qquad s_{i,t-1}^2 = \omega_i + a_i s_{i,t-2}^2 + b_i r_{i,t-1}^2 \tag{5.3}$$

for which $\boldsymbol{\gamma}_i = (\omega_i, a_i, b_i)$.

Consider the stochastic regression model (5.2). As noted in Section 3.2, a key ingredient in the optimal weight vector that solves the optimization problem (3.1) is $(\boldsymbol{\mu}_n, \mathbf{V}_n)$, where $\boldsymbol{\mu}_n = E(\mathbf{r}_{n+1}|\mathcal{R}_n)$ and $\mathbf{V}_n = E(\mathbf{r}_{n+1}\mathbf{r}_{n+1}^T|\mathcal{R}_n)$. Instead of the classical model of i.i.d. returns, one can combine domain knowledge of the m assets with time series modeling to obtain better predictors of future returns via $\boldsymbol{\mu}_n$ and $\mathbf{V}_n$. The regressors $\mathbf{x}_{i,t-1}$ in (5.2) can be chosen to build a combined substantive–empirical model for prediction; see Section 7.5 of Lai and Xing (2008). Since the model (5.2) is intended to produce i.i.d. $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \ldots, \epsilon_{mt})^T$, or i.i.d. $\mathbf{z}_t = (z_{1t}, \ldots, z_{mt})^T$ after adjusting for conditional heteroskedasticity as in (5.3), we can still use the NPEB approach to determine the optimal weight vector, bootstrapping from the estimated common distribution of $\boldsymbol{\epsilon}_t$ (or $\mathbf{z}_t$). Note that (5.2) and (5.3) models the asset returns separately, instead of jointly in a multivariate regression or multivariate GARCH model which has too many parameters to estimate. While the vectors $\boldsymbol{\epsilon}_t$ (or $\mathbf{z}_t$) are assumed to be i.i.d., (5.2) [or (5.3)] does not

assume their components to be uncorrelated since it treats the components separately rather than jointly. The conditional cross-sectional covariance between the returns of assets i and j given $\mathcal{R}_n$ is given by

$$(5.4) \quad \mathrm{Cov}(r_{i,n+1}, r_{j,n+1}|\mathcal{R}_n) = s_{i,n}(\boldsymbol{\gamma}_i) s_{j,n}(\boldsymbol{\gamma}_j) \, \mathrm{Cov}(z_{i,n+1}, z_{j,n+1}|\mathcal{R}_n)$$

for the model (5.2) and (5.3). Note that (5.3) determines $s_{i,n}^2$ recursively from $\mathcal{R}_n$, and that $\mathbf{z}_{n+1}$ is independent of $\mathcal{R}_n$ and, therefore, its covariance matrix can be consistently estimated from the residuals $\widehat{\mathbf{z}}_t$. Under (5.2) and (5.3), the NPEB approach uses the following formulas for $\boldsymbol{\mu}_n$ and $\mathbf{V}_n$ in (3.5):

$$(5.5) \quad \boldsymbol{\mu}_n = (\widehat{\boldsymbol{\beta}}_1^T \mathbf{x}_{1,n}, \ldots, \widehat{\boldsymbol{\beta}}_m^T \mathbf{x}_{m,n})^T, \qquad \mathbf{V}_n = \boldsymbol{\mu}_n \boldsymbol{\mu}_n^T + (\widehat{s}_{i,n} \widehat{s}_{j,n} \widehat{\sigma}_{ij})_{1 \leq i,j \leq n},$$

in which $\widehat{\boldsymbol{\beta}}_i$ is the least squares estimate of $\boldsymbol{\beta}_i$, and $\widehat{s}_{l,n}$ and $\widehat{\sigma}_{ij}$ are the usual estimates of $s_{l,n}$ and $\mathrm{Cov}(z_{i,1}, z_{j,1})$ based on $\mathcal{R}_n$. Further discussion of time series modeling for implementing the optimal portfolio in Section 3 will be given in Sections 6.2 and 7.

6. An empirical study. In this section we describe an empirical study of the out-of-sample performance of the proposed approach and other methods for mean–variance portfolio optimization when the means and covariances of the underlying asset returns are unknown. The study uses monthly stock market data from January 1985 to December 2009, which are obtained from the Center for Research in Security Prices (CRSP) database, and evaluates out-of-sample performance of different portfolios of these stocks for each month after the first ten years (120 months) of this period to accumulate training data. The CRSP database can be accessed through the Wharton Research Data Services at the University of Pennsylvania (http://wrds.wharton.upenn.edu). Following Ledoit and Wolf (2004), at the beginning of month t, with t varying from January 1995 to December 2009, we select $m = 50$ stocks with the largest market values among those that have no missing monthly prices in the previous 120 months, which are used as the training sample. The portfolios for month t to be considered are formed from these m stocks.

Note that this period contains highly volatile times in the stock market, such as around "Black Monday" in 1987, the Internet bubble burst and the September 11 terrorist attacks in 2001, and the "Great Recession" that began in 2007 with the default and other difficulties of subprime mortgage loans. We use sliding windows of $n = 120$ months of training data to construct portfolios of the stocks for the subsequent month. In contrast to the Black–Litterman approach described in Section 2.2, the portfolio construction is based solely on these data and uses no other information about the stocks and their associated firms, since the purpose of the empirical study is to illustrate the basic statistical aspects of the proposed method and

to compare it with other statistical methods for implementing Markowitz's mean–variance portfolio optimization theory. Moreover, for a fair comparison, we do not assume any prior distribution as in the Bayes approach, and only use NPEB in this study.

Performance of a portfolio is measured by the excess returns e_t over a benchmark portfolio. As t varies over the monthly test periods from January 1995 to December 2009, we can (i) add up the realized excess returns to give the *cumulative realized excess return* $\sum_{l=1}^{t} e_l$ up to time t, and (ii) use the average realized excess return and the standard deviation to evaluate the *realized information ratio* $\sqrt{12}\bar{e}/s_e$, where $\bar{e}$ is the sample average of the monthly excess returns and s_e is the corresponding sample standard deviation, using $\sqrt{12}$ to annualize the ratio as in Ledoit and Wolf (2004). Noting that the realized information ratio is a summary statistic of the monthly excess returns in the 180 test periods, we find it more informative to supplement this commonly used measure of investment performance with the time series plot of cumulative realized excess returns, from which the realized excess returns e_t can be retrieved by differencing.

We use two ways to construct the benchmark portfolio. The first follows that of Ledoit and Wolf (2004), who propose to mimic how an active portfolio manager chooses the benchmark to define excess returns. It is described in Section 6.1. The second simply uses the S&P500 Index as the benchmark portfolio and Section 6.3 considers this case. Section 6.2 compares the time series of the returns of these two benchmark portfolios and explains why we choose to use the S&P500 Index as the benchmark portfolio in conjunction with the time series model (5.2) and (5.3) for the excess returns in Section 6.3.

6.1. *Active portfolios and associated optimization problems.* In this section the benchmark portfolio consists of the $m = 50$ stocks chosen at the beginning of each test period and weights them by their market values. Let $\mathbf{w}_B$ denote the weight of this value-weighted benchmark and $\mathbf{w}$ the weight of a given portfolio. The difference $\widetilde{\mathbf{w}} = \mathbf{w} - \mathbf{w}_B$ satisfies $\widetilde{\mathbf{w}}^T \mathbf{1} = 0$. An active portfolio manager would choose $\mathbf{w}$ that solves the following optimization problem instead of (1.1):

$$\mathbf{w}_{\text{active}} = \mathbf{w}_B + \arg\min_{\widetilde{\mathbf{w}}} \widetilde{\mathbf{w}}^T \boldsymbol{\Sigma} \widetilde{\mathbf{w}} \quad \text{subject to} \quad \widetilde{\mathbf{w}}^T \boldsymbol{\mu} = \widetilde{\mu}_*,$$

(6.1)
$$\widetilde{\mathbf{w}}^T \mathbf{1} = 0 \text{ and } \widetilde{\mathbf{w}} \in \mathcal{C},$$

in which $\mathcal{C}$ represents additional constraints for the manager, $\boldsymbol{\Sigma}$ is the covariance matrix of stock returns and $\widetilde{\mu}_*$ is the target excess return over the value-weighted benchmark. The portfolio defined by $\mathbf{w}_{\text{active}}$ is called an *active portfolio*. Since $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are typically unknown, putting a prior distribution on them in (6.1) leads to the following modification of (3.1):

(6.2) $$\max\{E(\widetilde{\mathbf{w}}^T \mathbf{r}_{n+1}) - \lambda \operatorname{Var}(\widetilde{\mathbf{w}}^T \mathbf{r}_{n+1})\} \quad \text{subject to} \quad \widetilde{\mathbf{w}}^T \mathbf{1} = 0.$$

TABLE 3
Means and standard deviations (in parentheses) of the annualized realized excess returns over the value-based benchmark

$\widetilde{\mu}_*$ λ	0.01 2^2	0.015 2	0.02 2^{-1}	0.03 2^{-2}
(a) All test periods by re-defining portfolios in some periods				
Plug-in	0.001 (4.7e−3)	0.002 (7.3e−3)	0.003 (9.6e−3)	0.007 (1.4e−2)
Shrink	0.003 (4.3e−3)	0.004 (6.6e−3)	0.006 (8.8e−3)	0.011 (1.3e−2)
Boot	0.001 (2.5e−3)	0.001 (3.8e−3)	0.001 (5.1e−3)	0.003 (7.3e−3)
NPEB	0.029 (1.2e−1)	0.046 (1.3e−1)	0.053 (1.5e−1)	0.056 (1.6e−1)
(b) Test periods in which all portfolios are well defined				
Plug-in	0.002 (6.6e−3)	0.004 (1.0e−2)	0.006 (1.4e−2)	0.014 (1.9e−2)
Shrink	0.005 (5.9e−3)	0.008 (9.0e−3)	0.012 (1.2e−2)	0.021 (1.8e−2)
Boot	0.001 (3.5e−3)	0.003 (5.3e−3)	0.003 (7.1e−3)	0.006 (1.0e−2)
NPEB	0.282 (9.3e−2)	0.367 (1.1e−1)	0.438 (1.1e−1)	0.460 (1.1e−2)

This optimization problem can be solved by the same method as that introduced in Section 3.

Following Ledoit and Wolf (2004), we choose the constraint set $\mathcal{C}$ such that the portfolio is long only and the total position in any stock cannot exceed an upper bound c, that is, $\mathcal{C} = \{\widetilde{\mathbf{w}} : -\mathbf{w}_B \leq \widetilde{\mathbf{w}} \leq c\mathbf{1} - \mathbf{w}_B\}$, with $c = 0.1$. We use quadratic programming to solve the optimization problem (6.1) in which $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are replaced, for the plug-in active portfolio, by their sample estimates based on the training sample in the past 120 months. The covariance-shrinkage active portfolio uses a shrinkage estimator of $\boldsymbol{\Sigma}$ instead, shrinking toward a patterned matrix that assumes all pairwise correlations to be equal [Ledoit and Wolf (2003)]. Similarly, we can extend Section 2.3 to obtain a resampled active portfolio, and also extend the NPEB approach in Section 4 to construct the corresponding NPEB active portfolio. Table 3 summarizes the realized information ratio $\sqrt{12}\bar{e}/s_e$ for different values of annualized target excess returns $\widetilde{\mu}_*$ and "matching" values of λ whose choice is described below.

We first note that specified target returns $\widetilde{\mu}_*$ may be vacuous for the plug-in, covariance-shrinkage (abbreviated "shrink" in Table 3) and resampled (abbreviated "boot" for bootstrapping) active portfolios in a given test period. For $\widetilde{\mu}_* = 0.01, 0.015, 0.02, 0.03$, there are 92, 91, 91 and 80 test periods, respectively, for which (6.1) has solutions when $\boldsymbol{\Sigma}$ is replaced by either the sample covariance matrix or the Ledoit–Wolf shrinkage estimator of the training data from the previous 120 months. Higher levels of target returns result in even fewer of the 180 test periods for which (6.1) has solutions. On the other hand, values of $\widetilde{\mu}_*$ that are lower than 1% may be of little practical interest to active portfolio managers. When (6.1) does not have a solution

to provide a portfolio of a specified type for a test period, we use the value-weighted benchmark as the portfolio for the test period. Table 3(a) gives the actual (annualized) mean realized excess returns $12\bar{e}$ to show the extent to which they match the target value $\widetilde{\mu}_*$, and also the corresponding annualized standard deviations $\sqrt{12}s_e$, over the 180 test periods for the plug-in, covariance-shrinkage and resampled active portfolios constructed with the above modification. These numbers are very small, showing that the three portfolios differ little from the benchmark portfolio, so the realized information ratios that range from 0.24 to 0.83 for these active portfolios can be quite misleading if the actual mean excess returns are not taken into consideration.

We have also tried another default option that uses 10 stocks with the largest mean returns (among the 50 selected stocks) over the training period and puts equal weights to these 10 stocks to form a portfolio for the ensuing test period for which (6.1) does not have a solution. The mean realized excess returns $12\bar{e}$ when this default option is used are all negative (between -17.4% and -16.3%), while $\widetilde{\mu}_*$ ranges from 1% from 3%. Table 3(a) also gives the means and standard deviations of the annualized realized excess returns of the NPEB active portfolio for four values of λ that are chosen so that the mean realized excess returns roughly match the values of $\widetilde{\mu}_*$ over a grid of the form $\lambda = 2^j$ $(-2 \leq j \leq 2)$ that we have tried. Note that NPEB has considerably larger mean excess returns than the other three portfolios.

Table 3(b) restricts only to the 80–92 test periods in which the plug-in, covariance-shrinkage and resampled active portfolios are all well defined by (6.1) for $\widetilde{\mu}_* = 0.01, 0.015, 0.02$ and 0.03. The mean excess returns of the plug-in, covariance-shrinkage and resampled portfolios are still very small, while those of NPEB are much larger. The realized information ratios of NPEB range from 3.015 to 3.954, while those of the other three portfolios range from 0.335 to 1.214 when we restrict to these test periods.

6.2. *Value-weighted portfolio versus S&P500 Index and time series effects.* The results for the plug-in and covariance-shrinkage portfolios in Table 3 are markedly different from those of Ledoit and Wolf (2004) covering a different period (February 1983–December 2002). This suggests that the stock returns cannot be approximated by the assumed i.i.d. model underlying these methods. In Section 5.3 we have extended the NPEB approach to a very flexible time series model (5.2) and (5.3) of the stock returns r_{it}. The stochastic regression model (5.2) can incorporate important time-varying predictors in $\mathbf{x}_{it}$ for the ith stock's performance at time t, while the GARCH model (5.3) for the random disturbances ϵ_{it} in (5.2) can incorporate dynamic features of the stock's idosyncratic variability. It seems that a regressor such as the return u_t of S&P500 Index should be included in $\mathbf{x}_{it}$ to take advantage of the co-movements of r_{it} and u_t. However, since u_t is not observed at time t, one may need to have good predictors of u_t which should consist
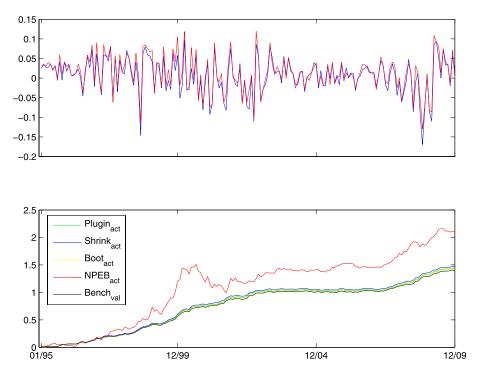
FIG. 2. *Time series plots of the monthly returns* (Top) *of the S&P500 Index and the value-weighted benchmark, and cumulative excess returns* (Bottom) *of the value-weighted benchmark (abbreviated by Bench*$_\mathrm{val}$*) and for active portfolios (abbreviated by the subscript "act") over the S&P500 Index.*

not only of the past S&P500 returns but also macroeconomic variables. Of course, stock-specific information such as the firm's earnings performance and forecast and its sector's economic outlook should also be considered. This means that fundamental analysis, as carried out by professional stock analysts and economists in investment banks, should be incorporated into the model (5.2). Since this is clearly beyond the scope of the present empirical study whose purpose is to illustrate our new statistical approach to the Markowitz optimization enigma, we shall focus on simple models to demonstrate the benefit of building good models for $\mathbf{r}_{t+1}$ in our stochastic optimization approach.

In this connection, we first compare the S&P500 Index with the value-weighted portfolio, which is the benchmark portfolio in Section 6.1. The top panel of Figure 2 gives the time series plots of the monthly returns (which are not annualized) of both portfolios during the test period. The S&P500 Index has mean 0.006 and standard deviation 0.046 in this period, while the mean of the value-weighted portfolio is 0.0137 and its standard deviation is 0.045. The bottom panel of Figure 2 plots the time series of
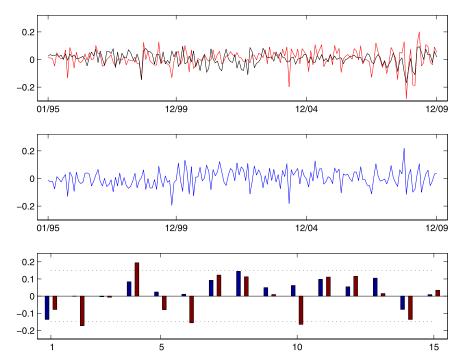
FIG. 3. *Comparison of returns and excess returns.* Top panel: *returns of S&P500 Index (black) and SHW (red).* Middle panel: *excess returns (blue) of SHW.* Bottom panel: *Autocorrelations of returns (red) and excess returns (blue) of SHW; the dotted lines represent rejection boundaries of 5%-level tests of zero autocorrelation at indicated lag.*

cumulative realized excess returns $\sum_{l=1}^{t} e_l$ over the S&P500 Index, for the value-weighted portfolio and also for the four active portfolios in Table 3(a) under the column $\mu_* = 0.015$ and $\lambda = 2$, during the test period (January 1995–December 2009). Unlike NPEB$_{\text{act}}$, the cumulative realized excess returns of the other three active portfolios differ little from the value-weighted portfolio, as shown by the figure.

In view of the structural changes in the economy and the financial markets during this period, it appears difficult to find simple time series models that can reflect the inherent nonstationarity. If we use the S&P500 Index u_t as an alternative benchmark to the value-weighted portfolio used in Section 6.1, the excess returns $e_{it} = r_{it} - u_t$ may be able to exploit the co-movements of r_{it} and u_t to remove their common nonstationarity due to changes in macroeconomic variables. As an illustration, the top panel of Figure 3 gives the time series plots of returns of Sherwin–Williams Co. (SHW) and of the S&P500 Index during this period, and the middle panel gives the time series plot of the excess returns. The Ljung–Box test, which involves autocorrelations of lags up to 20 months, has p-value 0.001 for the monthly returns

of SHW and 0.267 for the excess returns, and therefore rejects the i.i.d. assumption for the actual but not the excess returns; see Section 5.1 of Lai and Xing (2008). This is also shown graphically by the autocorrelation functions in the bottom panel of Figure 3.

6.3. *Using the S&P500 Index as benchmark portfolio and time series models of excess returns.*   The preceding section shows that using the S&P500 Index as the benchmark portfolio has certain advantages over the value-weighted portfolio. In this section we consider the excess returns e_{it} over the S&P500 Index u_t, which we use as the benchmark portfolio, and fit relatively simple time series models to the training sample to predict the mean and volatility of e_{it} for the test period. Instead of forming active portfolios as in Section 6.1, we follow traditional portfolio theory as described in Sections 1–3. Note that this theory assumes the constraint $\sum_{i=1}^{m} w_i = 1$ and, therefore,

$$(6.3) \qquad \sum_{i=1}^{m} w_i r_{it} - u_t = \sum_{i=1}^{m} w_i(r_{it} - u_t) = \sum_{i=1}^{m} w_i e_{it},$$

whereas active portfolio optimization considers weights $\widetilde{w}_i = w_i - w_{i,B}$ that satisfy the constraint $\sum_{i=1}^{m} \widetilde{w}_i = 0$. In view of (6.3), when the objective is to maximize the mean return of the portfolio subject to a constraint on the volatility of the excess return over the benchmark (which is related to achieving an optimal information ratio), we can replace the returns r_{it} by the excess returns e_{it} in the portfolio optimization problem (1.1) or (3.1). As explained in the second paragraph of Section 6.2, e_{it} can be modeled by simpler stationary time series models than r_{it}.

The simplest time series model to try is the AR(1) model $e_{it} = \alpha_i + \gamma_i e_{i,t-1} + \epsilon_{it}$. Assuming this time series model for the excess returns, we can apply the NPEB procedure in Section 5.3 to the training sample and thereby obtain the NPEB$_{\text{AR}}$ portfolio for the test sample. The AR(1) model uses $\mathbf{x}_{i,t-1} = (1, e_{i,t-1})^T$ as the predictor in a linear regression model for $e_{i,t}$. To improve prediction performance, one can include additional predictor variables, for example, the return u_{t-1} of the S&P500 Index in the preceding period. Assuming the stochastic regression model $e_{i,t} = (1, e_{i,t-1}, u_{t-1})\boldsymbol{\beta}_i + \epsilon_{i,t}$, and the GARCH(1, 1) model (5.3) for $\epsilon_{i,t}$, we can apply the NPEB procedure to the training sample and thereby form the NPEB$_{\text{SRG}}$ portfolio for the test sample.

Instead of taking long-only positions (i.e., $w_i \geq 0$ for all i), we also allow short selling, with the constraint $w_i \geq -0.05$ for all i, to construct the following portfolios in this section. For the plug-in, covariance-shrinkage and resampled portfolios, which we abbreviate as in Figure 2 but without the subscript "act" (for active), we use the annualized target return
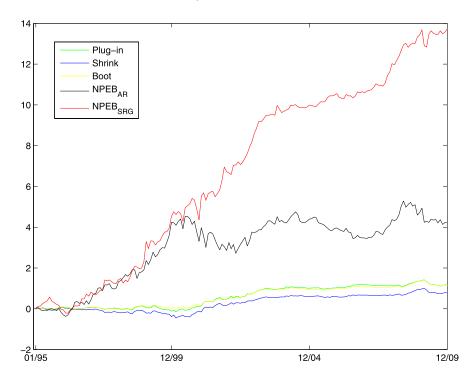
FIG. 4. *Realized cumulative excess returns over the S&P500 Index.*

$\mu_* = 0.015, 0.02, 0.03$, for which the problem (1.1) can be solved for all 180 test periods under the weight constraint; note that we use the mean return instead of the mean excess return as the target μ_*. For the NPEB$_{AR}$ and NPEB$_{SRG}$ portfolios, we use the training sample as in Section 5.1 to choose λ by maximizing the information ratio over the grid $\lambda \in \{2^i : i = -3, -2, \ldots, 6\}$. Figure 4 plots the time series of cumulative realized excess returns over the S&P500 Index during the test period of 180 months, for Plug-in, Shrink and Boot with $\mu_* = 0.015$ and for NPEB$_{AR}$ and NPEB$_{SRG}$. Table 4 gives the annualized realized information ratios, with the S&P500 Index as the benchmark portfolio. The table also considers cases $\mu_* = 0.02, 0.03$, and further abbreviates Plug-in, Shrink and Boot by P, S, B, respectively.

6.4. *Discussion.* Our approach may perform much better if the investor can combine domain knowledge with the statistical modeling that we illustrate here. We have not done this in the present comparative study because using a purely empirical analysis of the past returns of these stocks to build the prediction model (5.2) would be a disservice to the power and versatility of the proposed approach, which is developed in Section 3 in a general Bayesian framework, allowing the skillful investor to make use of prior beliefs on the future return vector $\mathbf{r}_{n+1}$ and statistical models for predicting $\mathbf{r}_{n+1}$

TABLE 4

Realized information ratios and average realized excess returns (in square brackets) with respect to the S&P500 Index

$\mu_* = 0.015$			$\mu_* = 0.02$			$\mu_* = 0.03$			NPEB	
P	S	B	P	S	B	P	S	B	AR	SRG
0.527	0.352	0.618	0.532	0.353	0.629	0.538	0.354	0.625	0.370	1.169
[0.078]	[0.052]	[0.077]	[0.078]	[0.051]	[0.078]	[0.076]	[0.050]	[0.077]	[0.283]	[0.915]

from past market data. The prior beliefs can involve both the investor's and the market's "views," as in the Black–Litterman approach described in Section 2.2, for which the market's view is implied by the equilibrium portfolio. Note that Black and Litterman model the potential errors of these views by normal priors whose covariance matrices reflect the uncertainties. Our Bayesian approach goes one step further to account for these uncertainties by using the *actual* means and variances of the portfolio's return in the optimization problem (3.1), instead of the *estimated* means and variances in the plug-in approach.

A portfolio on Markowitz's efficient frontier can be interpreted as a minimum-variance portfolio achieving a target mean return, or a maximum-mean portfolio at a given volatility (i.e., standard derivation of returns). Portfolio managers prefer the former interpretation, as target returns are appealing to investors. In active portfolio management [Grinold and Kahn (2000)], this has led to the target excess return $\widetilde{\mu}_*$ and the optimization problem (6.1). The empirical study in Section 6.1 shows that when the means and covariances of the stock returns are unknown and are estimated from historical data, putting these estimates in (6.1) may not provide a solution; moreover, the actual mean of the solution (when it exists) can differ substantially from $\widetilde{\mu}_*$.

7. Concluding remarks. The "Markowitz enigma" has been attributed to (a) sampling variability of the plug-in weights (hence use of resampling to correct for bias due to nonlinearity of the weights as a function of the mean vector and covariance matrix of the stocks) or (b) inherent difficulties of estimation of high-dimensional covariance matrices in the plug-in approach. Like the plug-in approach, subsequent refinements that attempt to address (a) or (b) still follow closely Markowitz's solution for efficient portfolios, constraining the unknown mean to equal to some target returns. This tends to result in relatively low information ratios when no or limited short selling is allowed, as noted in Sections 4.3 and 6. Another difficulty with the plug-in and shrinkage approaches is that their measure of "risk" does not account for the uncertainties in the parameter estimates. Incorporating these

uncertainties via a Bayesian approach results in a much harder stochastic optimization problem than Markowitz's deterministic optimization problem, which we have been able to solve by introducing an additional parameter η.

Our solution of this stochastic optimization problem opens up new possibilities in extending Markowitz's mean–variance portfolio optimization theory to the case where the means and covariances of the asset returns for the next investment period are unknown. As pointed out in Section 5.3, our solution only requires the posterior mean and second moment matrix of the return vector for the next period, and one can combine the Black–Litterman-type expert views with statistical modeling to develop Bayesian or empirical Bayes models with good predictive properties, for example, by using (5.2) with suitably chosen $\mathbf{x}_{i,t-1}$.

SUPPLEMENTARY MATERIAL

Supplement: Matlab implementation of the NPEB method
(DOI: 10.1214/10-AOAS422SUPP; .zip). The source code of our approach is provided.

REFERENCES

Best, M. J. and Grauer, R. R. (1991). On the sensitivity of mean–variance-efficient portfolios to changes in asset means: Some analytical and computational results. *Rev. Fin. Stud.* **4** 315–342.

Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969

Black, F. and Litterman, R. (1990). *Asset Allocation: Combining Investor Views with Market Equilibrium.* Goldman, Sachs and Co., New York.

Britten-Jones, M. (1999). The sampling error in estimates of mean–variance efficient portfolio weights. *J. Finance* **54** 655–671.

Broadie, M. (1993). Computing efficient frontiers using estimated parameters. *Ann. Oper. Res.* **45** (Special Issue on Financial Engineering) 21–58.

Canner, N., Mankiw, G. and Weil, D. N. (1997). An asset allocation puzzle. *Amer. Econ. Rev.* **87** 181–191.

Chopra, V. K., Hensel, C. R. and Turner, A. L. (1993). Massaging mean–variance inputs: Returns from alternative global investment strategies in the 1980s. *Management Sci.* **39** 845–855.

Fabozzi, F. J., Kolm, P. N., Pachamanova, P. A. and Focardi, S. M. (2007). *Robust Portfolio Optimization and Management.* Wiley, New York.

Fan, J., Fan, Y. and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *J. Econometrics* **147** 186–197. MR2472991

Frankfurter, G. M., Phillips, H. E. and Seagle, J. P. (1976). Performance of the Sharpe portfolio selection model: A comparison. *J. Financ. Quant. Anal.* **6** 191–204.

Grinold, R. C. and Kahn, R. N. (2000). *Active Portfolio Management*, 2nd ed. McGraw-Hill, New York.

HUANG, J. Z., LIU, N., POURAHMADI, M. and LIU, L. (2006). Covariance matrix selection and estimation via penalized normal likelihood. *Biometrika* **93** 85–98. MR2277742

JOBSON, J. D. and KORKIE, B. (1980). Estimation for Markowitz efficient portfolios. *J. Amer. Statist. Assoc.* **75** 544–554. MR0590686

JORION, P. (1986). Bayes–Stein estimation for portfolio analysis. *J. Financ. Quant. Anal.* **21** 279–292.

LAI, T. L. and XING, H. (2008). *Statistical Models and Methods for Financial Markets.* Springer, New York. MR2434025

LEDOIT, P. and WOLF, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empirical Finance* **10** 603–621.

LEDOIT, P. and WOLF, M. (2004). Honey, I shrunk the sample covariance matrix. *J. Portfolio Management* **30** 110–119.

MARKOWITZ, H. M. (1952). Portfolio selection. *J. Finance* **7** 77–91.

MARKOWITZ, H. M. (1959). *Portfolio Selection.* Wiley, New York. MR0103768

MEUCCI, A. (2005). *Risk and Asset Allocation.* Springer, New York. MR2155219

MEUCCI, A. (2010). The Black–Litterman approach: Original model and extensions. In *The Encyclopedia of Quantitative Finance* (R. Cont, ed.) **1** 196–199. Wiley, New York.

MICHAUD, R. O. (1989). *Efficient Asset Management.* Harvard Business School Press, Boston, MA.

PRESS, W. H., TEUKOLSKY, S. A., WETTERLING, W. T. and FLANNERY, B. P. (1992). *Numerical Recipes in C*, 2nd ed. Cambridge Univ. Press, Cambridge. MR1201159

ROSS, S. A. (1976). The arbitrage theory of capital asset pricing. *J. Econ. Theory* **13** 341–360. MR0429063

SHARPE, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *J. Finance* **19** 425–442.

SIMANN, Y. (1997). Estimation risk in portfolio selection: The mean variance model versus the mean absolute deviation model. *Management Sci.* **43** 1437–1446.

T. L. LAI
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
PALO ALTO, CALIFORNIA 94305
USA
E-MAIL: lait@stanford.edu

T. XING
DEPARTMENT OF APPLIED MATHEMATICS
    AND STATISTICS
STATE UNIVERSITY OF NEW YORK
    AT STONY BROOK
STONY BROOK, NEW YORK 11794
USA
E-MAIL: xing@ams.sunysb.edu

Z. CHEN
BOSERA FUNDS
SHENZHEN
CHINA
E-MAIL: zehao.chen@gmail.com