



Portfolio optimisation using constrained hierarchical bayes models¹

Jiangyong Yin and Xinyi Xu

Department of Statistics, The Ohio State University, Columbus, OH, USA

ABSTRACT

It is well known that traditional mean-variance optimal portfolio delivers rather erratic and unsatisfactory out-of-sample performance due to the neglect of estimation errors. Constrained solutions, such as no-short-sale-constrained and norm-constrained portfolios, can usually achieve much higher *ex post* Sharpe ratio. Bayesian methods have also been shown to be superior to traditional plug-in estimator by incorporating parameter uncertainty through prior distributions. In this paper, we develop an innovative method that induces priors directly on optimal portfolio weights and imposing constraints *a priori* in our hierarchical Bayes model. We show that such constructed portfolios are well diversified with superior out-of-sample performance. Our proposed model is tested on a number of Fama–French industry portfolios against the naïve diversification strategy and Chevrier and McCulloch's (2008) economically motivated prior (EMP) strategy. On average, our model outperforms Chevrier and McCulloch's (2008) EMP strategy by over 15% and outperform the '1/N' strategy by over 50%.

ARTICLE HISTORY

Received 20 April 2017
Accepted 22 June 2017

KEYWORDS

Bayesian hierarchical models;
parameter constrains;
portfolio choices

1. Introduction

Portfolio optimisation is a fundamental problem in financial research. Its most common formulation is Markowitz's (1952) mean-variance paradigm (see Brandt, 2009, for a detailed review). Suppose that an investor would like to choose a portfolio in the universe of a risk-free asset and p risky assets, which have returns r_f and r_t at time t , respectively, where r_t is a p -dimensional vector. Denote the excess returns by $R_t = r_t - r_f 1_p$, where 1_p is a vector of ones, and assume that $R_t \stackrel{iid}{\sim} N_p(\mu, \Sigma)$. Based on the observed returns R_1, \dots, R_T in that past T periods, the goal is to find an optimal portfolio weight vector w that minimises the utility function

$$\begin{aligned} U(w) &= w' E[R_{T+1}] - \frac{\gamma}{2} w' \text{Var}(R_{T+1}) w \\ &= w' \mu - \frac{\gamma}{2} w' \Sigma w, \end{aligned} \quad (1.1)$$

where R_{T+1} is the excess return vector to be realised in the next period and γ is the relative risk-aversion coefficient. If both μ and Σ were known, the optimal portfolio weight can be expressed in the closed-form

$$w^* = \arg \max_w U(w) = \frac{1}{\gamma} \Sigma^{-1} \mu, \quad (1.2)$$

and the corresponding utility is $U(w^*) = \mu' \Sigma^{-1} \mu / 2\gamma = \theta^2 / 2\gamma$, where $\theta^2 = \mu' \Sigma^{-1} \mu$ is the squared ratio of the *ex ante* tangency portfolio of the risky assets. However, in practice, the true values of μ and Σ are never known. To compute the portfolio weight, the traditional approach is to first estimate these parameters and then

plug the estimates into (1.2) as if they were the true parameters. This common practice, however, ignores the uncertainty in parameter estimation. The resulting portfolios usually have extreme long or short positions on very few assets, hence are not well diversified and have poor performances (see e.g. Bawa, Brown, & Klein, 1979).

Extensive efforts have been made to improve portfolio allocation, among which a very successful class of methods avoid extreme portfolio positions by effectively shrinking the weights towards a certain target. Under the Bayesian framework, this shrinkage can be achieved by placing prior distributions on the model parameters. Integrating out the parameters with respect to the posteriors leads to the predictive distribution $\pi(R_{T+1} | R_1, \dots, R_T)$ of the future outcome, and the optimal portfolio weight w is the Bayes rule that maximises the expected utility (1.1) under the predictive distribution. This approach naturally incorporates the uncertainty in parameter estimation, and thus may lead to sharper risk assessment and better decision-making (see e.g. Avramov & Zhou, 2010). However, the performance of a Bayesian procedure is impacted by the choice of the priors. Frost and Savarino (1986) showed that an empirical Bayes model based on informative priors could outperform a simple Bayesian model based on non-informative priors. Greyserman, Jones, and Strawderman (2006) showed that for a long enough investment horizon, a hierarchical Bayes model with conjugate priors outperforms those based on non-informative priors and James–Stein priors, in terms of both direct utility factors

CONTACT Xinyi Xu  xu.214@osu.edu

¹Present address: CapitalG, Inc.

and turnover rates. Tu and Zhou (2010) developed ‘economic-objective-based’ priors and showed that the resulting portfolio can achieve better out-of-sample Certainty-Equivalent Returns (CER) and utility gains. Moreover, Chevrier and McCulloch (2008) incorporated economic theory into their priors and obtained excellent out-of-sample Sharpe ratios and turnover rates.

Another natural approach for avoiding extreme positions is to impose pre-specified constraints on the portfolio weight w . The most commonly used constraints are the no-short-sale constraint (i.e., $w \geq 0$) and \mathcal{L}_1 -/ \mathcal{L}_2 -norm constraints (i.e. $|w| \leq c$ or $|w|^2 < c$, where c is a pre-specified constant). The no-short-sale constraint was first explored by Frost and Savarino (1988). Under the mean-variance paradigm, it has been shown that imposing the no-short-sale constraint can be viewed as shrinking the expected return towards the mean (DeMiguel, Garlappi, & Uppal, 2009b). Moreover, Jagannathan and Ma (2003) showed that under the alternative minimal variance paradigm, where

$$\begin{aligned} w &= \arg \min \tilde{U}(w) = \arg \min w' \Sigma w \\ \text{s.t. } w' \mathbf{1} &= 1, \end{aligned} \quad (1.3)$$

imposing the no-short-sale constraint is equivalent to shrinking the sample estimate of the covariance matrix, which is a well-studied method pursued by many researchers including Ledoit and Wolf (2003a) and Ledoit and Wolf (2003b). The \mathcal{L}_1 -/ \mathcal{L}_2 -norm constraints were carefully studied by DeMiguel, Garlappi, Nogales, and Uppal (2009a), who showed that under the minimal-variance paradigm, certain \mathcal{L}_1 -constraints lead to the no-short-sale-constrained solution and certain \mathcal{L}_2 -constraints are equivalent to that of Ledoit and Wolf (2003a). More recent developments along this line include the sparsity penalty method proposed by Brodie, Daubechies, De Mol, Giannone, and Loris (2009).

Inspired by the successes of the Bayesian approach and the constrained optimisation approach, in this paper, we propose an innovative method that restricts the parameter space over which hierarchical priors span. We demonstrate the superior performances of our method through a series of empirical studies, comparing with the traditional mean-variance plug-in strategy, the so-called ‘1/N’ or naïve diversification strategy, and Chevrier and McCulloch (2008)’s Economically Motivated Prior (EMP) strategy. The ‘1/N’ strategy, which simply assigns equal weights to all assets, is a common benchmark used in the literature. DeMiguel et al. (2009b) examined many competing portfolio optimisation methods and showed that no method can outperform the ‘1/N’ strategy consistently. As far as we know, Chevrier and McCulloch’s (2008) EMP strategy is one of the very few methods that have been shown to outperform the ‘1/N’ portfolio on a large collection of data-sets. Our investigation demonstrates that the

proposed restricted Bayesian strategy outperforms both the ‘1/N’ strategy and Chevrier and McCulloch (2008)’s EMP strategy under several measures.

The remainder of the paper is organised as follows. In Section 2, we first propose two hierarchical priors based on the hyper- g and the economic-objective-based priors, and then impose the no-short-sale constraint on the parameter space as *a priori*. In Section 3, we construct MCMC algorithms for fitting the proposed restricted Bayesian models. Then we evaluate the performance of the proposed restricted Bayesian models through a series of empirical analysis in Section 4. Finally, we summarise the findings and discuss future directions in Section 5.

2. Model specification

Under the mean-variance paradigm, the unknown parameters in the distribution of the excess returns are μ and Σ . Common Bayesian portfolio optimisation models can be represented by

$$\begin{aligned} R_1, \dots, R_T | \mu, \Sigma &\stackrel{\text{iid}}{\sim} N_p(\mu, \Sigma), \\ \mu | \tau, \Sigma &\sim N_p(\mu_0, \tau \Sigma), \\ \Sigma &\sim \pi(\Sigma), \end{aligned} \quad (2.1)$$

where μ_0 is the investor’s view of the mean asset returns, τ determines the strength of belief in the value of μ_0 , and $\pi(\Sigma)$ is usually a conjugate, or a non-informative, or a more complicated hierarchical prior on the covariance matrix Σ . Frost and Savarino (1986) utilised the empirical Bayes method to estimate μ_0 , τ and the hyper-parameters in an inverse-Wishart prior on Σ , and showed that the resulting portfolio overperforms that under the non-informative priors on both μ and Σ . Greyserman et al. (2006) further proposed a hierarchical prior that places another layer of priors on μ_0 and Σ , and showed that the fully hierarchical Bayes procedure produces promising results compared to the classic procedures.

As pointed out by Tu and Zhou (2010) and the references therein, traditional diffuse priors on μ and Σ could imply very informative and unreasonable priors on w . To avoid this pitfall, it is better to use the first-order condition (1.2) to reparameterise the model (2.1), so that the priors are placed directly on w and Σ rather than on μ and Σ . The same approach has been used in Kandel, McCulloch, and Stambaugh (1995) and Lamoureux and Zhou (1996), and a similar idea has been used in the famous Black–Litterman model (Black & Litterman, 1992). Replacing μ by $\mu = \gamma \Sigma w$ based on the first-order condition, we obtain the following reparameterised model:

$$\begin{aligned} R_1, \dots, R_T | w, \Sigma &\stackrel{\text{iid}}{\sim} N_p(\gamma \Sigma w, \Sigma), \\ w | \tau, \Sigma &\sim N_p\left(\frac{1}{\gamma} \Sigma^{-1} \mu_0, \frac{1}{\gamma^2} \tau \Sigma^{-1}\right), \\ \Sigma &\sim \pi(\Sigma). \end{aligned} \quad (2.2)$$

It is interesting to notice that conditional on Σ , the excess return R_t can be viewed to follow a classic normal linear regression model with the design matrix $X = \gamma \Sigma$ and the regression coefficient w . There is a large literature on prior elicitation for linear regression coefficients (see e.g. Clyde & George, 2004). One of the most widely adopted classes of priors are Zellner's (1986) g -priors, where w follows a multivariate normal distribution with the covariance matrix $g(X' \Sigma X)^{-1}$ and Σ follows a non-informative prior (Liang, Paulo, Molina, Clyde, & Berger, 2008). In this portfolio optimisation problem, note that $(X' \Sigma X)^{-1} = \Sigma^{-1}/\gamma^2$. Therefore, (2.2) can be viewed as placing a g -prior on the portfolio weight w , where $g = \tau$.

Conditional on τ and Σ , the posterior mean of w can be represented by

$$E[w|\mathbf{R}, \tau, \Sigma] = \frac{1}{1 + \tau T} \frac{1}{\gamma} \Sigma^{-1} \mu_0 + \frac{\tau T}{1 + \tau T} \frac{1}{\gamma} \Sigma^{-1} \bar{R}_T,$$

where \bar{R}_T is the sample average of the excess returns, that is, the optimal portfolio weight is a weighted average of the prior portfolio weight $\frac{1}{\gamma} \Sigma^{-1} \mu_0$ and the 'plug-in' estimator $\frac{1}{\gamma} \Sigma^{-1} \bar{R}_T$. The prior parameter τ , scaled by the length of the observation window T , adjusts the shrinkage degree of the plug-in estimator towards the prior portfolio weights, and hence plays an important role in this optimisation problem. To decide the value of τ , Greyserman et al. (2006) fixed τ at a pre-specified value. However, as shown by Liang et al. (2008), fixed choices of τ do not utilise the information in the data, and thus may cause undesirable consistency issues for model selection. Instead, they recommended integrating the marginal likelihood under a proper prior on τ , and showed that under certain regularity conditions, such mixtures of g -priors resolve many problems with the fixed τ priors, while maintaining the computational tractability. In this paper, we follow Liang et al.'s (2008) suggestion and use the hyper- g priors with

$$p(\tau|\alpha) \propto (1 + \tau T)^{-\alpha}, \quad \alpha > 0. \quad (2.3)$$

This family of priors includes the priors used by Strawderman (1971) to provide improved mean square risk over ordinary maximum-likelihood estimates in the normal means problem. These priors have also been studied by Cui and George (2008) for the problem of variable selection in the case of known error variance. We further extend this class of priors by placing a conjugate prior $\pi(\alpha) \propto e^{-\alpha}$ on the hyper-parameter α . As in (mixtures of) g priors, we place a non-informative prior $\pi(\Sigma) \propto |\Sigma|^{-\frac{p+1}{2}}$ on the covariance matrix Σ .

Furthermore, note that in the model (2.2), the portfolio weight w follows a multivariate normal distribution conditional on τ and Σ , which assigns positive prior weight to any non-empty open set in the \mathcal{R}^p space. Therefore, the portfolio weight potentially can still take extreme long or short positions. To alleviate this danger

and to future stabilise the portfolio choices, we investigate placing regularisations or constraints on the portfolio weight. Such regularisations have been shown to be very helpful for improving out-of-sample performances for financial returns data which typically has very low signal-to-noise ratio (see e.g. Brodie et al., 2009; DeMiguel et al., 2009a, 2009b; Jagannathan & Ma, 2003). Under the Bayesian framework, the parameters are considered as random variables, and so the ranges of their values are determined by the supports of their distributions. The reparameterisation (2.2) models w directly, and thus allows us to place constraint on the portfolio weight by restricting the support of the prior on w to a subset in \mathcal{R}^p . As mentioned in Section 1, a widely used constraint in the literature is the no-short-sale constraint $w \geq 0$. Combining this constraint with the Bayesian hierarchical model (2.2) yields the following *restricted Return-based Hierarchical Bayes* (rRHB) model:

$$\begin{aligned} R_1, \dots, R_T | w, \Sigma &\stackrel{\text{iid}}{\sim} N_p(\gamma \Sigma w, \Sigma), \\ w | \tau, \Sigma &\sim N_p\left(\frac{1}{\gamma} \Sigma^{-1} \mu_0, \frac{1}{\gamma^2} \tau \Sigma^{-1}\right) \\ &\quad \times I(w \in [0, \infty)^p), \\ \pi(\tau | \alpha) &\propto (1 + \tau T)^{-\alpha}, \\ \pi(\alpha) &\propto e^{-\alpha}, \\ \pi(\Sigma) &\propto |\Sigma|^{-\frac{p+1}{2}}. \end{aligned} \quad (2.4)$$

The first-order condition (1.2) holds for any pair of parameter values μ and Σ , and so also holds on this restricted subspace.

It is important to note that our cRHB model (2.4) is fundamentally different from the constrained optimisation approaches in many previous works, including Jagannathan and Ma (2003), DeMiguel et al. (2009a), DeMiguel et al. (2009b) and Brodie et al. (2009). We are simply restricting the parameter space so that the optimal portfolio weights w is non-negative for all assets, but an individual investor is still allowed to hold short positions on any asset as he/she wants (i.e. $E(w|R_1, \dots, R_T)$ could have negative components based on certain sets of observations). On the other hand, the constrained optimisation approaches restrict all investors' action spaces by imposing constraints on the portfolio weights that they can choose, but do not constrain the parameter space. Therefore, the first-order condition (1.2) does not hold for these constrained optimisation solutions under either the mean variance or minimal variance framework, but still holds for our cRHB model. To the best of our knowledge, the only other paper that has similarly constrained the parameter space is Chevrier and McCulloch (2008), but their underlying hierarchical model is entirely different from ours. We will compare our method with that in Chevrier and McCulloch (2008) through empirical studies in Section 4.

3. Model fitting

We use the Gibbs sampler to simulate from the joint posterior distributions of the proposed Bayesian hierarchical models. The algorithm for a single iteration in the Gibbs sampler is described below.

- *Step 1* – Update w from the truncated multivariate normal distribution $N(\mu_w, \Sigma_w)I\{w \in (0, \infty)^p\}$, where

$$\mu_w = \frac{\tau T}{1 + \tau T} \frac{1}{\gamma} \Sigma^{-1} \bar{R}_T + \frac{1}{1 + \tau T} \frac{1}{\gamma} \Sigma^{-1} \mu_0,$$

$$\Sigma_w = \frac{\tau}{\gamma^2(1 + \tau T)} \Sigma^{-1}.$$

By the results in Rodriguez-Yam, Davis, and Scharf (2004), sampling from the truncated multivariate normal distribution can be implemented by sampling from a series of truncated univariate normal distributions

$$w_{[i]} | w_{[-i]} \sim N(\mu_{[i]}, \Sigma_{[i]}) I\{w_{[i]} \geq 0\},$$

$$i = 1, \dots, p$$

where $\mu_{[i]} = \mu_{[i]} + \Sigma_{[i,-i]} \Sigma_{[-i,-i]}^{-1} (w_{[-i]} - \mu_{[-i]})$,

$$\Sigma_{[i]} = \Sigma_{[i,i]} - \Sigma_{[i,-i]} \Sigma_{[-i,-i]}^{-1} \Sigma_{[-i,i]}.$$

- *Step 2* – Update τ through a Metropolis–Hastings algorithm, where the conditional posterior distribution is

$$p(\tau | \mathbf{R}, \alpha, \Sigma) \propto \tau^{-p/2} (1 + \tau T)^{-\alpha}$$

$$\exp \left\{ -\frac{\gamma^2}{2\tau} \left(w - \frac{1}{\gamma} \Sigma^{-1} \eta_0 \right)' \Sigma \left(w - \frac{1}{\gamma} \Sigma^{-1} \eta_0 \right) \right\},$$

and the proposal distribution is a truncated normal distribution on the positive values.

- *Step 3* – Update α from the exponential distribution with mean $(1 + \log(1 + \tau T))^{-1}$.
- *Step 4* – Update Σ from its conditional posterior distribution

$$p(\Sigma | \mathbf{R}, \tau, w) \propto |\Sigma|^{-(T+p)/2}$$

$$\exp \left\{ -\frac{1}{2} \text{trace}(\Sigma^{-1} \Psi + \Sigma \Phi) \right\},$$

where $\Psi = \sum_{t=1}^T R_t R_t' + \frac{1}{\tau} \eta_0 \eta_0'$ and $\Phi = \gamma^2 (T + \frac{1}{\tau}) w w'$. This distribution is a multivariate version of the famous generalised inverse Gaussian (GIG) distribution, and is sometimes called the matrix GIG distribution (Butler, 1998). Sampling from this distribution is much harder than sampling from the common Wishart or inverse Wishart distributions. Most of the existing sampling algorithms depend on some form of matrix decomposition. In this paper, we use the hit-and-run sampler designed by Yang and Berger (1994), since it works reasonably well for the sizes of matrix considered in our empirical studies. This sampler can be described as follows:

- (1) Generate a random lower triangle matrix L with elements $l_{ij} \stackrel{\text{iid}}{\sim} N(0, 1)$ for $i \leq j$.
- (2) Define a step matrix D by scaling and reflecting L . Let $s = \sqrt{\sum_{(i,j): i \leq j} l_{ij}^2}$, then

$$\tilde{L} = \frac{1}{s} L, \quad \text{and} \quad D = \tilde{L} + \tilde{L}^T - \text{diag}(\tilde{L}).$$

- (3) Generate a random step-size variable $\lambda \sim N(0, 1)$.
- (4) Given the last update $\Sigma^{[k]}$, propose a new update

$$\Sigma^{\text{prop}} = \exp(\log(\Sigma^{[k]}) + \lambda D),$$

where $\exp(\cdot)$ and $\log(\cdot)$ are matrix functions defined by $\exp(\mathbf{A}) = \sum_{m=0}^{\infty} \mathbf{A}^m / m!$.

- (5) Accept Σ^{prop} with probability

$$\min \left(1, \frac{p(\Sigma^{\text{prop}} | \mathbf{R}, \tau, w)}{p(\Sigma^{[k]} | \mathbf{R}, \tau, w)} \right).$$

Otherwise, set $\Sigma^{[k+1]} = \Sigma^{[k]}$.

For the other analysis, different samplers for the matrix GIG distribution could be implemented for this step as well.

Note that in the above algorithm, the values of w are directly drawn from its conditional posterior distribution, and thus the optimal portfolio weight can be simply computed as the average of these posterior draws.

4. Empirical studies

We now evaluate the performances of our proposed cRHB model (2.4) through comparisons with the following benchmark methods in the literature:

- *MeanVar*: The traditional mean-variance plug-in portfolio. It simply plugs the sample mean and sample covariance matrix into the first-order condition (1.2), which yields

$$w = \frac{\hat{\Sigma}^{-1} \hat{\mu}}{\mathbf{1}' \hat{\Sigma}^{-1} \hat{\mu}}.$$

- *1/N*: The naïve diversification strategy. When there are p assets in the portfolio, it assigns each asset exactly the same weight, i.e.

$$w = \frac{1}{p} \mathbf{1}_p.$$

Despite its simple form, this naïve diversification strategy has excellent performances. DeMiguel et al. (2009b) performed a comprehensive analysis comparing it with a large collection of competing methods, including the mean variance plug-in method, the minimal variance optimisation under no-short-sale constraints, the James–Stein shrinkage estimator, etc. They found that none of these

methods can consistently and significantly outperform this naïve diversification strategy in terms of out-of-sample Sharpe ratio and CER. As a result, a lot of recent research works evaluate new methods by comparing with only the 1/N portfolio.

- *EMP*: The Bayesian portfolio using economic theory in Chevrier and McCulloch (2008). It is under a hierarchical prior constructed from the CAMP model, the no-short-sale constraint, and another constraint that guarantees that the market portfolio lies on the efficient part of the mean variance frontier. The authors compared their portfolio with the “1/N” strategy on 27 domestic and international data-sets, and discovered that their approach overperforms by over 30% on average in terms of out-of-sample Sharpe ratio, while maintaining a similar low turnover. As far as we know, this is one of the very few strategies in the literature that have been shown to outperform the 1/N portfolio on a large collection of data-sets.

For the cRHB model, we construct the portfolios using relative weights $\frac{w}{|w|_1}$ rather than absolute weights w . The parameter μ_0 that represents the investor’s view of the mean asset returns in cRHB is set to be a vector of zeros.

4.1. Data

To facilitate the comparison, we use the following data-sets from DeMiguel et al. (2009b) and Chevrier and McCulloch (2008) – the Fama French 5, 10 and 49 industry portfolios and the 6 and 25 portfolios formed on size and book-to-market; and also an additional data-set – the Fama French 17 industry portfolio. All the data were retrieved from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. The data-sets on this website are regularly updated and span nearly 80 years. In order to allow direct comparison with the results of Chevrier and McCulloch (2008), we use the data in the same time period as theirs – from August 1963 to July 2007.

Following Chevrier and McCulloch (2008), we prepare the data by subtracting the risk-free rate from the portfolio returns and augmenting the data with the Fama French 3 factors. Monthly market excess returns are used for the industry assets, and nominal returns are used for the Fama and French factor assets. We also remove the industry labelled ‘other’ from the data when considering the Fama French industry portfolios, since in theory the industry portfolios span the market.

To evaluate the out-of-sample performances, as in Ledoit and Wolf (2003b) and Chevrier and McCulloch (2008), we use 10 years of data (from August of year $t - 10$ to July of year t , $T = 120$) to obtain the portfolio weights. The formed portfolios are then held from the

first day in August of year t to the last day in July of year $t + 1$. We then repeat this ‘rolling window’ procedure, i.e. we assume automatic yearly rebalance of all portfolio assets. In total, we obtain 34 years of out-of-sample performance based on the realised excess returns from August 1973 to July 2007.

4.2. Out-of-sample performance criteria

The out-of-sample performance of a portfolio can be measured under many different criteria. For example, Greyserman et al. (2006) examined the Sharpe ratio, portfolio turnover and the utility gain. Chevrier and McCulloch (2008) used the Sharpe ratio and turnover. DeMiguel et al. (2009b) compared the Sharpe ratio, turnover and certainty-equivalent return (CER), while Tu and Zhou (2010) mainly examined CER and the out-of-sample utility gain. To facilitate our comparison with the ‘1/N’ and the ‘EMP’ strategies, we use the Sharpe ratio and turnover, as suggested by Chevrier and McCulloch (2008) and DeMiguel et al. (2009b).

4.2.1. Sharpe ratio

This is probably the most common portfolio performance measure in the literature. By the definition of Sharpe (1994), the single-period *ex post* Sharpe ratio compared to risk-free asset is

$$SR_1 = \frac{\hat{\mu}}{\hat{\sigma}}, \quad (4.1)$$

where $\hat{\mu}$ is the mean of out-of-sample series of one-month excess returns of the portfolio for all 34 test periods, and $\hat{\sigma}$ is the corresponding standard deviation. In practice, annualised Sharpe ratio is typically reported. For monthly return data, the annualised Sharpe ratio is given by $\sqrt{12}SR_1$.

4.2.2. Turnover

The turnover provides an indication of the trading volume for a particular strategy, as well as an upper bound for the transactions costs that such a strategy would entail. Following DeMiguel et al. (2009b) and Chevrier and McCulloch (2008), we let turnover be the sum of the absolute value of the rebalancing trades across the p available assets and over the T trading dates, normalised by the total number of trading dates T , that is,

$$\frac{1}{T} \sum_{t=1}^T \sum_{i=1}^p |w_t^i - w_{t-}^i|, \quad (4.2)$$

where w_t^i is the weight invested in asset i at time t , and w_{t-}^i is the weight invested in asset i just prior to rebalancing, which can be calculated as

$$w_{t-}^i = \frac{w_t^i(1 + R_t^i)}{\sum_j w_t^j(1 + R_t^j)}.$$

Table 1. Annualised Sharpe ratio.

Portfolios	Ind5	Ind10	Ind17	Ind49	Size6	Size25
MeanVar	65.91	64.41	55.07	41.93	27.19	55.25
1/N	56.41	54.71	49.05	48.24	58.24	57.10
EMP	72.7	83.7	—*	58.9	69.0	71.3
cRHB	89.82	83.94	73.47	74.34	90.33	81.03

*Not included in Chevrier and McCulloch (2008).

Table 2. Average monthly turnover.

Portfolios	Ind5	Ind10	Ind17	Ind49	Size6	Size25
MeanVar	116.87	133.95	169.59	541.92	95.32	3862.63
1/N	9.32	9.65	10.04	12.24	8.69	7.38
EMP	11.04	10.92	—*	11.64	8.88	9.12
cRHB	10.12	10.74	11.61	17.03	9.41	10.13

*Not included in Chevrier and McCulloch (2008).

Therefore, the turnover (4.2) can be interpreted as the average percentage of wealth traded at each time when we rebalance the portfolio.

4.3. Findings

Tables 1 and 2 summarise the Sharpe ratios and the turnovers of the five competing strategies, respectively. We highlight the highest Sharpe ratio for each data-set by boldface. It is easy to see that the cRHB model consistently has the highest Sharpe ratio, the Bayesian portfolio EMP usually has the second highest Sharpe ratio, and not surprisingly, the traditional mean-variance plug-in portfolio usually has the lowest Sharpe ratio. On average, the cRHB strategy overperforms the EMP strategy by 19%. For all the data-sets that we consider here, EMP overperforms the ‘1/N’ strategy, which is consistent with the results in Chevrier and McCulloch

(2008). The cRHB strategy overperforms the ‘1/N’ strategy by as high as 53%. As for turnover, the cRHB strategy has turnovers only slightly larger than those of the ‘1/N’ strategy, and comparable to those of the EMP strategy. The traditional mean-variance plug-in portfolio has extremely large turnovers, and thus could yield high transaction costs.

It is worth noting that across the data-sets of various sizes that we consider, the advantages of the cRHB strategy in Sharpe ratio remain quite stable. One of the claims in DeMiguel et al. (2009b) is that as the number of available assets increases, the ‘1/N’ strategy should perform increasingly better relative to other methods as its portfolio weight has no estimation error. However, for the data-sets that we examine, the Sharpe ratio improvements of the cRHB strategy over the ‘1/N’ strategy are between 42% and 55% with no obvious patterns. This shows that this restricted Bayesian strategy is robust and performs well even when the data are scarce. To better understand the properties of the cRHB strategy relative to those of the ‘1/N’ strategy, we demonstrate the means of the standard deviations of their out-of-sample excess returns in Table 3 (we do not include the EMP strategy here, because it is computationally very expensive to run the MCMC for their Bayesian model. All the numbers of the EMP strategy are based on the reports in Chevrier and McCulloch (2008)). It is clear to see that the ‘1/N’ strategy has the highest returns. However, it also has the highest standard deviations. The cRHB strategy is relatively more conservative with lower standard deviations at small prices of sacrificing a little bit return. Combining these two factors together, the cRHB strategy has higher overall Sharpe ratios.

Table 3. Annualised portfolio excess return and standard deviation.

Portfolios	Ind5	Ind10	Ind17	Ind49	Size6	Size25
MeanVar	3.93 (5.96)	4.21 (6.54)	3.37 (6.12)	4.72 (11.3)	0.21 (0.75)	15.01 (27.17)
1/N	6.06 (10.7)	7.52 (12.9)	6.95 (14.2)	7.68 (15.9)	7.52 (12.92)	9.25 (16.14)
cRHB	4.85 (5.40)	5.21 (6.21)	5.46 (7.44)	6.26 (8.42)	5.85 (6.48)	7.12 (8.78)

Table 4. Annualized Sharpe ratios in five-year periods between 1973 and 2007.

Time periods	Portfolios	Ind5	Ind10	Ind17	Ind49	Size6	Size25
08/73–07/07	cRHB	89.82	83.94	73.47	74.34	90.33	81.03
	1/N	56.41	54.71	49.05	48.23	58.24	57.10
08/73–07/78	cRHB	83.99	64.55	50.61	50.44	95.94	72.90
	1/N	18.83	15.00	15.68	19.93	47.93	47.47
08/78–07/83	cRHB	81.93	76.71	66.36	65.65	85.81	74.35
	1/N	61.48	57.46	49.32	55.66	72.04	69.70
08/83–07/88	cRHB	104.33	114.67	106.37	108.22	101.93	93.35
	1/N	46.93	53.16	44.87	34.12	35.34	33.49
08/88–07/93	cRHB	91.35	94.86	87.10	84.50	76.86	79.85
	1/N	67.52	70.34	60.07	53.43	57.35	56.98
08/93–07/98	cRHB	184.15	158.93	136.07	144.42	145.60	129.55
	1/N	152.80	142.76	114.10	114.92	116.58	104.27
08/98–07/03	cRHB	17.72	15.38	17.12	20.23	39.90	47.74
	1/N	13.50	4.58	8.74	13.08	28.05	32.31
08/03–07/07	cRHB	155.85	162.42	155.33	160.64	156.08	150.71
	1/N	124.54	137.51	136.21	135.65	117.19	113.07

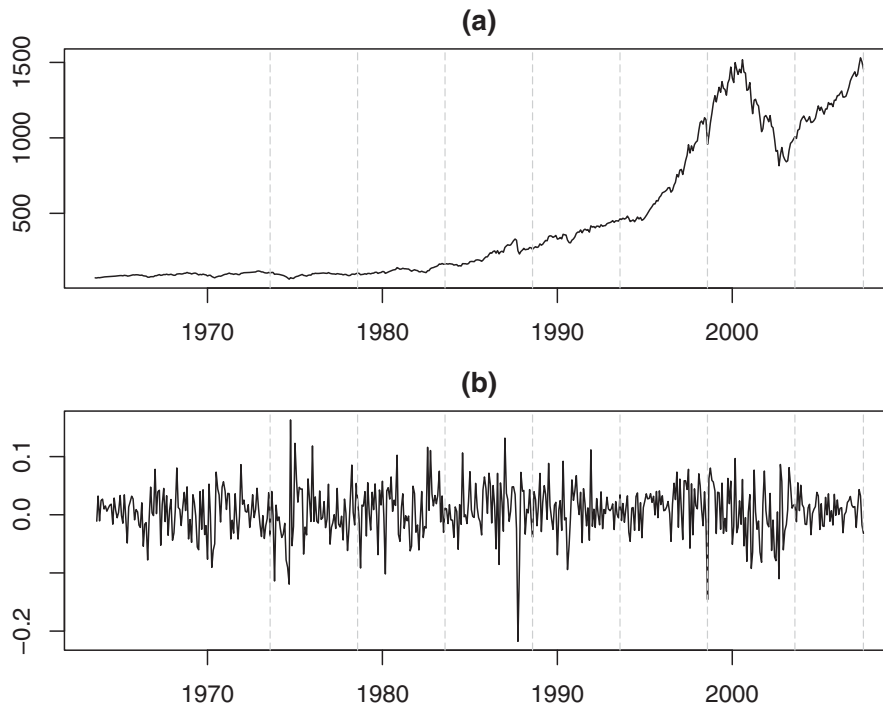


Figure 1. SP500 monthly closing price (a) and return (b) from August 1963 to July 2007.

Furthermore, it is easy to see from Definition (4.2.1) that the Sharpe ratio of a portfolio may fluctuate through time. To make sure that the cRHB strategy consistently has superior performance through time and to further examine its properties in different market environments, we follow the approach of Brodie et al. (2009) to calculate the Sharpe ratios by five-year periods (and four years for the last period). The results are shown in Table 4. It can be seen that the cRHB strategy overperforms the ‘1/N’ strategy in all these five-year periods and has the biggest advantages during three periods: August 1973–July 1978, August 1983–July 1988 and August 1998–July 2003 with the latter two covering the Black Monday in 1987 as well as the burst of dot-com bubble. The August 1973–July 1978 period seems relatively stable in comparison with recent decades, but we can see an increase in volatility in the return series (Figure 1(b)) which represents the 1973–1974 stock market crash. The ‘1/N’ strategy, due to the large standard deviations of its excess returns, is highly risky and could yield large negative returns during recessions and financial crises. On the other hand, the proposed cRHB strategy provides great balances between the excess returns and their standard deviations, and thus has outstanding performance over all the time periods.

5. Discussion

In this paper, we propose an innovative restricted Bayesian model, the cRHB model, which combines the advantages of the Bayesian approach and the

constrained optimisation approach. This model is robust and automatic in the sense that it allows subject inputs only through the choice of the prior center. We show through empirical studies that this model on average outperforms Chevrier and McCulloch’s (2008) EMP strategy by over 15% and outperforms the ‘1/N’ strategy by over 50%.

Note that in the cRHB model, the investor’s view of the mean asset returns μ_0 can be elicited from experts’ opinions or historical information. However, Tu and Zhou (2010) argued that before observing any data or does any formal statistical analysis, investors might have much better ideas about the optimal portfolio weight w rather than the values of μ . For example, the Black–Litterman model suggests using the value-weighted market portfolio weights, and the ‘1/N’ strategy suggests using equal portfolio weights for investment diversity. Therefore, an alternative approach could be to simply replace the prior mean of w by $w_0 = (1/p, \dots, 1/p)'$, which leads to the following *restricted Weight-based Hierarchical Bayes* (cWHB) model

$$\begin{aligned}
 R_1, \dots, R_T | w, \Sigma &\stackrel{\text{iid}}{\sim} N_p(\gamma \Sigma w, \Sigma) \\
 w | \tau, \Sigma &\sim N_p \left(w_0, \frac{1}{\gamma^2} \tau \Sigma^{-1} \right) I(w \in [0, \infty)^p) \\
 p(\tau | \alpha) &\propto (1 + \tau T)^{-\alpha} \\
 p(\alpha) &\propto e^{-\alpha} \\
 p(\Sigma) &\propto |\Sigma|^{-\frac{p+1}{2}}, \tag{5.1}
 \end{aligned}$$

where investors can choose either of the cRHB or cWHB models based on their knowledge of the mean asset returns or of the optimal portfolio weight.

Moreover, although both of our models use the no-short-sale constraint, the general methodology can be easily extended to accommodate other \mathcal{L}_1 -/ \mathcal{L}_2 -norm constraints in DeMiguel et al. (2009a). Also, in addition to the hyper-g prior (2.3) that we place on the shrinkage parameter τ , there exists a class of global / local shrinkage prior, such as the inverse gamma prior, the double exponential prior, the Strawderman–Berger prior, the horseshoe prior, etc. (see Carvalho, Polson, & Scott, 2010, for an insightful discussion). Some of these priors could lead to sparse strategies as the one in Brodie et al. (2009). It will be interesting to investigate the properties of the portfolio choices under these different priors.

Our experiences based on the empirical studies suggest that the performance of investment strategies depend on the investment period and the market environment. Therefore, instead of trying to find a single best optimisation strategy that works for all time periods, a more reasonable approach seems to be looking for different optimal solutions under various market conditions. Furthermore, the Bayesian framework allows us to incorporate the uncertainty of both the parameters and the models through Bayesian model averaging. It would be promising to construct optimal portfolios by taking weight averages of the portfolio weights from a collection of models, where the weights are determined by the posterior model probabilities.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported in part by US National Science Foundation (NSF) under grant DMS-1613110.

Notes on contributors

Jiangyong Yin received his PhD degree in statistics from The Ohio State University, working under the supervision of Dr. Xinyi Xu and Dr. Peter Craigmile. He is currently working at CapitalG Inc.

Xinyi Xu received his MS and PhD degrees in statistics from the University of Pennsylvania in 2003 and 2005, respectively. Currently, she is an associate professor in the Department of Statistics, The Ohio State University. Her research interests include Bayesian analysis, statistical decision theory, and high-dimensional data analysis.

References

- Avramov, D., & Zhou, G. (2010). Bayesian portfolio analysis. *Annual Review of Financial Economics*, 2(1), 25–47.
- Bawa, V., Brown, S., & Klein, R. (1979). *Estimation risk and optimal portfolio choice* (p. 190). New York: North Holland.
- Black, F., & Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48, 28–43.
- Brandt, M. (2009). Portfolio choice problems. In Y. Ait-Sahalia & L. P. Hansen (Eds.), *Handbook of financial econometrics: Tools and techniques* (pp. 269–336). Amsterdam: North-Holland.
- Brodie, J., Daubechies, I., De Mol, C., Giannone, D., & Loris, I. (2009). Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences*, 106(30), 12267–12272.
- Butler, R. W. (1998). Generalized inverse gaussian distributions and their wishart connections. *Scandinavian Journal of Statistics*, 25(1), 69–75.
- Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2), 465–480.
- Chevrier, T., & McCulloch, R. (2008). *Using economic theory to build optimal portfolios* (Technical report, Working paper). University of Chicago. Retrieved from <http://dx.doi.org/10.2139/ssrn.1126596>
- Clyde, M., & George, E. I. (2004). Model uncertainty. *Statistical Science*, 19, 81–94.
- Cui, W., & George, E. I. (2008). Empirical Bayes vs. fully Bayes variable selection. *Journal of Statistical Planning and Inference*, 138, 888–900.
- DeMiguel, V., Garlappi, L., Nogales, F., & Uppal, R. (2009a). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5), 798–812.
- DeMiguel, V., Garlappi, L., & Uppal, R. (2009b). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies*, 22(5), 1915.
- Frost, P., & Savarino, J. (1986). An empirical Bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis*, 21(03), 293–305.
- Frost, P., & Savarino, J. (1988). For better performance. *The Journal of Portfolio Management*, 15(1), 29–34.
- Greyserman, A., Jones, D., & Strawderman, W. (2006). Portfolio selection using hierarchical Bayesian analysis and MCMC methods. *Journal of Banking & Finance*, 30(2), 669–678.
- Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4), 1651–1684.
- Kandel, S., McCulloch, R., & Stambaugh, R. (1995). Bayesian inference and portfolio efficiency. *Review of Financial Studies*, 8(1), 1–53.
- Lamoureux, C., & Zhou, G. (1996). Temporary components of stock returns: What do the data tell us? *Review of Financial Studies*, 9(4), 1033–1059.
- Ledoit, O., & Wolf, M. (2003a). *Honey, I shrunk the sample covariance matrix* (UPF Economics and Business Working Paper 691). Retrieved from <http://dx.doi.org/10.2139/ssrn.433840>
- Ledoit, O., & Wolf, M. (2003b). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5), 603–621.
- Liang, F., Paulo, R., Molina, G., Clyde, M. A., & Berger, J. O. (2008). Mixtures of g priors for bayesian variable selection. *Journal of the American Statistical Association*, 103(481), 410–423.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77–91.
- Rodriguez-Yam, G., Davis, R., & Scharf, L. (2004). Efficient gibbs sampling of truncated multivariate normal with

- application to constrained linear regression. Unpublished Manuscript.
- Sharpe, W. (1994). The sharpe ratio. *Journal of Portfolio Management*, 21, 49–58.
- Strawderman, W. (1971). Proper bayes minimax estimators of the multivariate normal mean. *The Annals of Mathematical Statistics*, 42, 385–388.
- Tu, J., & Zhou, G. (2010). Incorporating economic objectives into Bayesian priors: Portfolio choice under parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 45(4), 959–986.
- Yang, R., & Berger, J. O. (1994). Estimation of a covariance matrix using the reference prior. *The Annals of Statistics*, 22, 1195–1211.