

Optimal Portfolio Diversification Using Maximum Entropy Principle¹

Anil K. Bera

Department of Economics

University of Illinois at Urbana-Champaign.

abera@uiuc.edu.

Sung Y. Park

WISE

Xiamen University.

sungpark@sungpark.net.

Forthcoming *Econometric Reviews*

¹We are grateful to the editors Amos Golan and Essie Maasoumi, two anonymous referees and Ximing Wu for many pertinent comments and helpful suggestions. In particular, Amos Golan's suggestion prompted us to use the generalized cross entropy measure (Section 4). We would also like to thank the participants of the Second Conference on Recent Developments in the Theory, Method and Applications of Information and Entropy Econometrics at the American University, Washington, D.C., September 23-25, 2005 for helpful comments and discussions. We, however, retain the responsibility for any remaining errors. Correspondence: Anil K. Bera, Department of Economics, University of Illinois at Urbana-Champaign, 1206 S. Sixth St, Champaign, IL 61820, U.S.A; Email: abera@uiuc.edu.

Abstract

Markowitz's mean-variance (MV) efficient portfolio selection is the one of the most widely used approaches in solving portfolio diversification problem. However, contrary to the notion of diversification, MV approach often leads to portfolios highly concentrated on a few assets. Also, this method leads to poor out-of-sample performances. Entropy is a well known measure of diversity and also has a shrinkage interpretation. In this paper, we propose to use cross entropy measure as the objective function with side conditions coming from the mean and variance-covariance matrix of the resampled asset returns. This automatically captures the degree of imprecision of input estimates. Our approach can be viewed as a shrinkage estimation of portfolio weights (probabilities) which are shrunk towards the predetermined portfolio, for example, equally weighted portfolio or minimum variance portfolio. Our procedure is illustrated with an application to the international equity indexes.

Key Words: Portfolio selection; Entropy measure; Shrinkage rule; Diversification; Simulation methods.

JEL Classification: C15; C44; G11.

1 Introduction

Markowitz's (1952) mean-variance (MV) optimization is one of the most common formulation of portfolio selection problem. However, portfolios constructed from sample moments of stock returns have proved problematic. The main problems in optimal MV portfolio are that the portfolios are often extremely concentrated on a few asset, which is a contradiction to the notion of diversification, and the out-of-sample performances of the MV portfolios are not very good. It is generally thought that these drawbacks are due to statistical error in estimating the moments that are used as inputs in the MV optimization. These errors are known to change optimal portfolio weights dramatically in such a way that portfolios often involve extreme positions (Jobson and Korkie, 1980). There have been extensive research on reducing statistical errors in sample mean and covariance matrix. One alternative is the class of shrinkage estimators. Frost and Savarino (1986), Jorion (1986), and Ledoit and Wolf (2003) used shrinkage estimation for the mean and covariance matrix. Shrinkage estimators compensate for the positive (negative) error that tends to be embedded in extremely high (low) estimated coefficients by pulling them downward (upward) and prevent extreme positions in portfolio selection.

Since shrinkage estimators are based on the empirical Bayesian approaches a particular prior distribution should be assumed to derive those estimators. Although some prior distributions used in the empirical Bayes estimation are known to work well, there is no systematic way to choose a prior distribution. For example, Jorion (1986) used an informative conjugate prior and derived the multivariate normal predictive distribution with the mean of minimum variance portfolio as the target mean. Frost and Savarino (1986) adapted a normal-wishart conjugate prior and derived multivariate Student's t predictive density. In their simulation study, they assumed that means, variances and correlations for all the assets are the same, so that their target mean and covariance matrix are those of equally weighted portfolio. As a result, it is very hard to achieve a certain shrinkage target preferred by asset managers, for example, a capitalization-weighted portfolio.

We propose a method that ensures shrinkage towards maximum diversification of portfolio weights using a information theoretic approach. Our objective function, the Kullback-Leibler information criteria (Kullback and Leibler, 1951) (KLIC) is defined as pseudo distance between two probability distributions (portfolio weights), $\mathbf{p} = (p_1, p_2, \dots, p_N)'$ and $\mathbf{q} = (q_1, q_2, \dots, q_N)'$:

$$KLIC(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^N p_i \ln(p_i/q_i). \quad (1)$$

The KLIC is also known as the cross-entropy (CE) measure (Golan, Judge, and Miller, 1996, p.29). If one minimizes the CE measure with \mathbf{q} as the reference distribution that satisfies certain constraints, one can get a solution, $\hat{\mathbf{p}}$, closest to \mathbf{q} . If we set $\mathbf{q} = (1/N, 1/N, \dots, 1/N)'$, uniform distribution, then $KLIC(\mathbf{p}, \mathbf{q})$ is same as negative Shannon's (1948) entropy measure. Since maximizing Shannon's entropy subject to some moment constraints implies estimating \mathbf{p} that is the closest to the uniform distribution (i.e., equally weighted portfolio), well-diversified optimal portfolio can be achieved.

In order to incorporate problems of imprecision of sample moments estimates, we define the confidence interval of maximized expected utility values which lead to inequality constraints to our optimization procedures. This confidence interval can be interpreted as the degree of uncertainty for the sample moments estimates, and can be estimated by resampling methods such as bootstrap or Monte-Carlo approaches.

There are several advantages in our information theoretic approach: (i) While previous papers primarily dealt with shrinkage estimators for the mean and covariance matrix to obtain more well-behaved optimal portfolios, we directly shrink portfolio composition(\mathbf{p}) towards pre-determined target portfolio weights(\mathbf{q}) that are of interest to asset managers; (ii) Most asset managers are not allowed to sell short (i.e., the portfolio weights cannot be negative) in the real world. Since constructed portfolio weights obtain through the maximum entropy (ME) approach are in the form of "probabilities," the weights are certainly non-negative. However, negative portfolio weights, when they are appropriate, for example, in case of hedge funds, can also be obtained using the generalized cross entropy (GCE) framework; (iii) Since the mean and covariance matrix should be estimated, one usually has only partial information. It is known that if sample sizes of individual returns are not large enough compare to the number of stocks, sample covariance matrix tends to be very imprecise. By minimizing the CE (or GCE) measure subject to certain well defined constraints, one can extract useful information from the sample mean and covariance matrix.

The rest of the paper is organized as follows. In Section 2, we provide a critical review of the existing methodologies. In Section 3, we discuss portfolio selection procedures using the ME principle based on the CE measure. In Section 4, the GCE formalism is proposed to obtain negative portfolio weights when short-selling is allowed. To illustrate the usefulness of our proposed methodologies, in Section 5, we provide an empirical application using eight international equity indexes with twelve different asset allocation models. The paper is concluded in Section 6.

2 Current Approaches To Portfolio Selection

We denote the first two moments of the excess returns $R = (R_1, R_2, \dots, R_N)' = (r_1 - r_f, r_2 - r_f, \dots, r_N - r_f)'$ on N risky assets as $\mathbb{E}(R) = (m_1, m_2, \dots, m_N)' = m$, and $Var(R) = ((\sigma_{ij})) = \Sigma$, a $N \times N$ matrix, where r_i and r_f denote the return of the i -th, $i = 1, 2, \dots, N$ and the risk-free assets, respectively. A portfolio $\pi = (\pi_1, \pi_2, \dots, \pi_N)'$ is a vector of weights that represents the investor's relative allocation of the wealth satisfying $\sum_{i=1}^N \pi_i = \pi' \mathbf{1}_N = 1$, where $\mathbf{1}_N$ is an $N \times 1$ vector of ones. The mean-variance (MV) problem is to choose the portfolio weight vector π to minimize the variance of the portfolio return $Var(\pi'R) = \pi' \Sigma \pi$ subject to a pre-determined target, μ_0 as expected return of the portfolio, i.e.,

$$\min_{\pi} \pi' \Sigma \pi, \quad \text{s.t.} \quad \mathbb{E}(\pi'R) = \pi' m = \mu_0, \quad \pi' \mathbf{1}_N = 1. \quad (2)$$

Merton (1972) obtained the Lagrange multipliers corresponding to the two constraints in (2), respectively, as

$$\gamma = \frac{C\mu_0 - A}{D}, \quad \nu = \frac{B - A\mu_0}{D},$$

where $A = \mathbf{1}'_N \Sigma^{-1} m$, $B = m' \Sigma^{-1} m$, $C = \mathbf{1}'_N \Sigma^{-1} \mathbf{1}_N$, and $D = BC - A^2$. The solution to (2) is given by

$$\hat{\pi} = \left(\frac{\mu_0}{B} \right) \Sigma^{-1} m$$

at which we have the MV portfolio variance as

$$\sigma_{\hat{\pi}}^2 = \hat{\pi}' \Sigma \hat{\pi} = \frac{C\mu_0^2 - 2A\mu_0 + B}{D}.$$

Therefore, we can write

$$\left(\frac{D}{C} \right) \sigma_{\hat{\pi}}^2 - \left(\mu_0 - \frac{A}{C} \right)^2 = \frac{D}{C^2}. \quad (3)$$

For a given mean and covariance matrix, the MV paradigm provide a very elegant way to achieve an efficient allocation such that higher expected returns can only be achieved by taking on more risk, as it is clear from the efficient frontier equation (3). Since the MV portfolio $\hat{\pi}$ is derived assuming investor's trade-off between the mean and the variance, the MV portfolio can also be obtained from the following expected utility maximization problem:

$$\max_{\pi} \mathbb{E}(\pi'R) - \frac{\lambda}{2} Var(\pi'R) \quad \text{s.t.} \quad \pi' \mathbf{1}_N = 1, \quad (4)$$

where λ denotes investor's degree of relative risk aversion.

There are, however, some drawbacks of the above MV paradigm. *First*, it is well known that the MV solution is very sensitive to estimation errors of mean m and covariance matrix Σ . Jobson and Korkie (1980) and Best and Grauer (1991) showed that the estimators such as the sample mean and sample covariance do not lend themselves to making inference in small sample, and small increase in the mean of just one asset drives half the securities out of the portfolio. *Second*, out-of-sample performance of the MV portfolio is very poor, as Jorion (1985) and DeMiguel, Garlappi, and Uppal (2005) showed, it is often even worse than the naive, equally weighted portfolio. *Finally*, related to the first point above, the MV optimal portfolio often has extreme portfolio weights due to statistical errors in mean and covariance estimates, which contradicts the notion of diversification. Michaud (1989) introduced the concept of "error maximization" because MV optimization overweight (underweight) those securities that have large (small) estimated returns, negative (positive) correlation and small (large) variance. To resolve these problems, a number of alternative methodologies have been proposed; some of which are discussed below.

2.1 Bayes-Stein shrinkage estimation

Suppose that the $(N \times 1)$ return vector R from N assets at time t ($t = 1, 2, \dots, T$) follows an IID multivariate normal distribution with mean μ and covariance matrix Σ , and the investor has an informative conjugate prior for μ

$$p(\mu|\bar{\mu}, \eta) \propto \exp \left[-\frac{1}{2}(\mu - \mathbf{1}_N \bar{\mu})'(\eta \Sigma^{-1})(\mu - \mathbf{1}_N \bar{\mu}) \right],$$

where $\bar{\mu}$ and η denote grand mean and prior precision, respectively. Then, the predictive density function of the vector of future return rate R^f , $p(R^f|R, \Sigma, \eta)$, is multivariate normal with predictive Bayes-Stein mean

$$\mu_{bs} = (1 - \phi_{bs})\hat{\mu} + \phi_{bs}\mu_{min}\mathbf{1}_N,$$

where $\hat{\mu}$ and μ_{min} denote the sample mean and the mean of minimum variance portfolio, respectively, and $\phi_{bs} = \eta/(T + \eta)$. Jorion (1986) adapted empirical Bayes-Stein estimation in the sense that he estimated the prior precision parameter, η , from the data assuming a gamma density for η with mean $(N+2)/(\mu - \mathbf{1}_N \bar{\mu})'\Sigma^{-1}(\mu - \mathbf{1}_N \bar{\mu})$. The shrinkage coefficient is estimated by

$$\phi_{bs} = \left(\frac{\eta}{T + \eta} \right) = \frac{N + 2}{(N + 2) + T(\hat{\mu} - \mu_{min}\mathbf{1}_N)'\hat{\Sigma}^{-1}(\hat{\mu} - \mu_{min}\mathbf{1}_N)}, \quad (5)$$

where $\hat{\Sigma}$ is the sample covariance matrix. Note that the non-informative Bayes-Stein estimator is a special case of (5) when $\eta = 0$ such that its mean and variance are given by $\hat{\mu}$ and $(1+1/T)\hat{\Sigma}$, respectively (Zellner and Chetty, 1965; Bawa, Brown and Klein, 1979). In this case, the sample mean is the predictive mean but the covariance matrix is inflated by $(1+1/T)$. Jorion (1986)'s method provides a reasonable strategy when investor's degree of belief about the estimated sample mean is weak. In the similar way, Ledoit and Wolf (2004a, 2004b) proposed shrinkage estimation for the covariance matrix Σ as

$$\hat{\Sigma}_{bs} = \delta \hat{F} + (1 - \delta) \hat{\Sigma},$$

where \hat{F} is usually chosen as a highly structured shrinkage target estimate. Ledoit and Wolf (2003) suggested the single-factor matrix of Sharpe (1963) as the shrinkage target and showed that their method substantially increases the realized information ratio of the portfolio manager. Frost and Savarino (1986) proposed shrinkage estimators for the mean and covariance at the same time. However, they assumed the same priors for all means, variances, and correlations, and thus the resulting portfolio weights shrunk towards the equally-weighted portfolio.

2.2 Imposing specific constraints

Frost and Savarino (1988) showed that imposing upper bound constraints or disallowing short-selling constraints on security weights reduces estimation bias and improves portfolio performance. On the other hand, Green and Hollifield (1992) argued that portfolio constraints may arrest the portfolio performance because some of the off-diagonal elements of Σ can take large negative values. Jagannathan and Ma (2003) showed that even if Green and Hollifield's argument is right, imposing non-negative constraints always helps, and has the same effect of using shrinkage estimate of Σ . Since shrinkage estimation improves finite sample behavior, imposing non-negative constraints also improves the portfolio performance.

2.3 Resampling approach

Resampling scheme enable us to evaluate how much MV optimized portfolio weights are affected by the error in estimating m and Σ . By drawing T observations B times without replacement from the empirical distribution using bootstrap, we obtain B new sets of the sample means and the sample covariance matrices $\{(\hat{m}_i, \hat{\Sigma}_i), i = 1, 2, \dots, B\}$. For each $(\hat{m}_i, \hat{\Sigma}_i)$, we get a sequence of optimized portfolio weights $\pi^i = (\pi_1^i, \pi_2^i, \dots, \pi_N^i)'$, $i = 1, 2, \dots, B$, by solving the MV problem or, equivalently,

maximizing the expected quadratic utility function. Evaluating (π^1, \dots, π^B) with the original inputs $(\hat{m}, \hat{\Sigma})$, we have B points of $\{(\hat{m}_{\pi^i}, \hat{\sigma}_{\pi^i}), i = 1, 2, \dots, B\}$, where $\hat{m}_{\pi^i} = \pi^{i'} \hat{m}$ and $\hat{\sigma}_{\pi^i} = \sqrt{\pi^{i'} \hat{\Sigma} \pi^i}$. These B points are statistically equivalent to the MV optimal efficient portfolio under the original inputs $(\hat{m}, \hat{\Sigma})$, and must lie below its frontier.

Michaud (1998) proposed resampled efficient portfolio using resampling method. Instead of considering a particular MV portfolio as above, let us consider MV portfolios on the MV efficient frontiers. By setting ranks for each MV efficient frontier between minimum variance portfolio (say, rank 1) and maximum return portfolio (say, rank l), B sets of rank-associated MV efficient portfolios can be calculated using $\{(\hat{m}_i, \hat{\Sigma}_i), i = 1, 2, \dots, B\}$ at each rank, $k = 1, \dots, l$, i.e., we have B portfolios for each rank. The resampled weight for a portfolio of rank k is given by

$$\bar{\pi}_k^{rs} = \frac{1}{B} \sum_{b=1}^B \pi_{b,k}, \quad (6)$$

where $\pi_{b,k}$ denotes the $N \times 1$ vector of rank- k portfolio for b -th resampling. The main difference between methods of the resampled efficient portfolio and the empirical Bayes portfolio is that in the former, we first do the optimization and then calculate final portfolio weights, while in the later optimization procedure is carried out at the second stage after obtaining the empirical Bayes-Stein estimates of m and Σ . Since the resampled weights are calculated by sample average of B number of resampling portfolios, it is well-diversified. However, Scherer (2002) pointed out that the distribution of weights, $\pi_{b,k}$ for $b = 1, \dots, B$ is usually skewed so that the sample mean cannot represent the location of the distribution correctly. In the next section, we propose our entropy approach to optimal portfolio selection which has nice interpretations of portfolio diversification and shrinkage effects.

3 Information theoretic approach to portfolio selection

3.1 Entropy measures

A discrete probability distribution $\mathbf{p} = (p_1, p_2, \dots, p_N)'$ of a random variable taking N values provides a measure of uncertainty (disorder) regarding that random variable. In the information theory literature, this measure of disorder is called entropy. Entropy measures have been extensively used in econometrics, and for more on this see, Maasoumi (1993), Golan, Judge and Miller (1996), Ullah (1996) and Bera and Biliias (2002).

A portfolio allocation $\pi = (\pi_1, \pi_2, \dots, \pi_N)'$ among N risky assets, with properties $\pi_i \geq 0$, $i = 1, 2, \dots, N$ and $\sum_{i=1}^N \pi_i = 1$, has the structure of a proper probability distribution. We will use the Shannon entropy (SE) measure

$$\text{SE}(\pi) = - \sum_{i=1}^N \pi_i \ln \pi_i \quad (7)$$

as a measure of portfolio diversification. When $\pi_i = 1/N$ for all i , $\text{SE}(\pi)$ has its maximum value $\ln N$. The other extreme case occurs when $\pi_i = 1$ for one i , and $= 0$ for the rest, then $\text{SE}(\pi) = 0$. Therefore, SE that provides a good measure of disorder in a system or expected information in a probability distribution, can be taken as a measure of portfolio diversification. In financial applications, portfolios are generally evaluated in terms of their degree of diversification using the SE measure after portfolios are obtained using different selection procedures (see for instance, Hoskisson, Hitt, Johnson and Moesel (1993), Lubatkin, Merchant and Srinivasan (1993) and Fernholz (2002, p.36)). We put the entropy itself in the objective function so as to obtain maximum diversity in a portfolio allocation. It is clear that when we maximize $\text{SE}(\pi)$ we shrink the portfolio towards an equally weighted portfolio, namely, $N^{-1}\mathbf{1} = (1/N, 1/N, \dots, 1/N)'$. We will also consider a more general objective function. Suppose a portfolio weight changes from π_i to q_i , then the change in entropy is $-\ln q_i - (-\ln \pi_i) = \ln(\pi_i/q_i)$. Taking average of $\ln(\pi_i/q_i)$ with π_i 's as weights we end up with the notion of cross-entropy (CE), $\text{CE}(\pi, \mathbf{q}) = \text{KLIC}(\pi, \mathbf{q})$, defined in (1). It is clear that when $\mathbf{q} = (1/N, 1/N, \dots, 1/N)'$, $\text{CE}(\pi, \mathbf{q}) = \sum_{i=1}^N \pi_i \ln \pi_i - \ln N$. Therefore, maximization of SE in (7) is a special case of CE minimization with respect to an equally weighted portfolio. In our analysis we will emphasize the minimization of $\text{CE}(\pi, \mathbf{q})$ for a given \mathbf{q} as a reasonable opportunity set for an investor. Thus, starting from an initial portfolio allocation \mathbf{q} , through minimization of CE we can obtain a more diversified portfolio. Golan, Judge and Miller (1996, p.31) showed that

$$\text{CE}(\pi, \mathbf{q}) = \sum_{i=1}^N \pi_i \ln(\pi_i/q_i) \approx \sum_{i=1}^N \frac{1}{q_i} (\pi_i - q_i)^2 \quad \text{for } q_i > 0. \quad (8)$$

Thus, we adjust small allocations of the initial portfolio \mathbf{q} more than the large ones, possibly resulting in a more diversified portfolio.

3.2 Preliminary approach

A good starting point for incorporating entropy measure in the portfolio selection is the dice problem introduced by Jaynes (1963). The dice problem can be stated as follows: Suppose one is asked to estimate the probabilities $\pi = (\pi_1, \pi_2, \dots, \pi_6)'$

for each possible outcomes of a fair six-sided die. The only information available is the mean value of the distribution, say μ_0 . There are infinite number of sets of values of π that will lead to the mean value of μ_0 . Jaynes (1963, p.187) suggested the need for a measure of the “uncertainty” of the probability distribution that can be maximized subject to the mean constraint which represents the available information, and advocated that a correct measure of uncertainty is the SE given in (7). As we mentioned before, portfolio weights for different financial assets can be regarded as probabilities: weights are non-negative and they sum to 1. Thus, we can consider portfolio selection problem such that asset managers are asked to select portfolio weights $\pi = (\pi_1, \pi_2, \dots, \pi_N)'$ for N assets conditional on a given investor’s preferred mean value of the portfolio, say μ_0 . This problem, like that of Jaynes’ is ill-posed since N number of weights need to be determined with only two pieces of information: mean of portfolio is equal to μ_0 and the sum of weights is equal to 1. Following Jaynes (1963) we can state the optimization problem as [see also Golan, Judge and Miller (1996, pp.12-14)]

$$\max_{\{\pi_i\}_{i=1}^N} - \sum_{i=1}^N \pi_i \ln \pi_i \quad (9)$$

subject to

$$\sum_{i=1}^N \hat{m}_i \pi_i = \mu_0, \quad \sum_{i=1}^N \pi_i = 1, \quad (10)$$

where \hat{m}_i denotes sample mean of asset i . After setting the Lagrangian function as

$$\mathfrak{L} = - \sum_{i=1}^N \pi_i \ln \pi_i - \gamma \left(\sum_{i=1}^N \hat{m}_i \pi_i - \mu_0 \right) - \lambda \left(\sum_{i=1}^N \pi_i - 1 \right),$$

we get the solution

$$\hat{\pi}_i = \frac{1}{\Omega(\gamma)} \exp[-\gamma \hat{m}_i], \quad i = 1, 2, \dots, N, \quad (11)$$

where $\Omega(\gamma) = \sum_{i=1}^N \exp[-\gamma \hat{m}_i]$ obtained by satisfying $\sum_{i=1}^N \pi_i = 1$.

The solution (11) turns out to be a probability mass function that has the form of an exponential distribution and therefore, it naturally yields no short-selling ($\hat{\pi}_i \geq 0$). Since the objective function (9) is same as the negative of $\text{CE}(\pi, \mathbf{q})$ with $\mathbf{q} = N^{-1}\mathbf{1}$ plus a constant, we can interpret the solution $\hat{\pi}$ as closest to the equally weighted portfolio (i.e., the most diversified portfolio) conditional on prescribed target mean μ_0 . In this sense, resulting portfolio weights are maximum diversified portfolio given mean constraint. However, this formulation uses information of return (mean) with-

out considering risk (variance). By including additional side constraint on variance $\sigma^2 = \pi' \Sigma \pi$, one can extend the above optimization problem as

$$\max_{\pi} -\pi' \ln \pi \quad (12)$$

subject to

$$\pi' \hat{m} \geq \mu^0, \quad \sqrt{\pi' \hat{\Sigma} \pi} \leq \sigma^0, \quad \pi \geq 0, \quad \text{and} \quad \pi' \mathbf{1}_N = 1, \quad (13)$$

where $\hat{\Sigma}$ denotes the sample covariance matrix of asset returns. The inequality constraints (13) can be interpreted as boundary conditions which an investor might prefer, i.e., the portfolio mean is not less than μ^0 and the portfolio standard deviation is not greater than σ^0 . Although the problem (12)-(13) is intuitively simple, it does not have a simple solution, primarily due to the nonlinear inequality constraint, $\sqrt{\pi' \Sigma \pi} \leq \sigma^0$.

Suppose that an investor is concerned with only mean ($\mu_{\pi} = \pi' m$) and standard deviation ($\sigma_{\pi} = \sqrt{\pi' \Sigma \pi}$) of portfolio returns. Then, one way to represent the inequality constraints in (13) is by the indifference curve of the Leontief utility function $U(\sigma^0, \mu^0)$. We can define investor i 's opportunity set due to the constraints in (13) by $\Xi_i = \{(\sigma_{\pi}, \mu_{\pi}) | \sigma_{\pi} \leq \sigma_i^0, \mu_{\pi} \geq \mu_i^0\}$. Suppose investors 'A' and 'B' choose particular lower bounds, μ_A^0 and μ_B^0 , and upper bounds, σ_A^0 and σ_B^0 , for portfolio means and standard deviations, respectively. In Figure 1, U_A and U_B denote two investors' indifference curves, and point E corresponds to the equally weighted portfolio. For each investor i , the maximization problem given in (12)-(13) is the same as choosing the closest portfolio weights to the equally weighted portfolio with $(\sigma_{\pi}, \mu_{\pi}) \in \Xi_i$.

[Figure 1]

By generating many possible values of m and Σ , we found numerically that portfolios which solve (12)-(13) lie on the vertical line of the indifference curve if $\sigma_i^0 < \sigma_E^0$ and $\mu_i^0 < \mu_E^0$ (i.e., investor 'A') and on the horizontal line of the indifference curve if $\sigma_i^0 > \sigma_E^0$ and $\mu_i^0 > \mu_E^0$. When $\sigma_i^0 < \sigma_E^0$ and $\mu_i^0 > \mu_E^0$ (i.e., 'B' investor), the portfolio solves above optimization problem at the kinked-point i (point B in the case of 'B' investor). In the case of $\Xi_E \subseteq \Xi_i$ (i.e., when $\sigma_i^0 > \sigma_E^0$ and $\mu_i^0 < \mu_E^0$), the maximum diversified portfolio is the equally weighted portfolio. Since the Leontief utility function is not differentiable it is hard to solve this problem by standard gradient-based optimization routines. Moreover, this model cannot account for estimation imprecision such as when we use the sample mean and covariance.

3.3 General approach

To incorporate estimation imprecision of the mean and covariance (as in Bayes-Stein estimation), we need more general constraint than in (13). In general, we consider the following minimization problem

$$\min_{\pi} \text{CE}(\pi|\mathbf{q}) = \sum_{i=1}^N \pi_i \ln(\pi_i/q_i) \quad (14)$$

subject to

$$\mathbb{E}U(\pi, R, \lambda) \geq \tau, \quad \pi \geq 0, \quad \text{and} \quad \pi' \mathbf{1}_N = 1, \quad (15)$$

where $U(\pi, R, \lambda)$ is an utility function, λ is the risk aversion parameter, and τ reflects investor's strength of belief in the estimated expected utility values, which we elaborate further below. We assume that $N \times 1$ random vector R has a distribution function $F(R)$ with density $f(R)$. To see the significance of τ , we define

$$\xi \equiv \mathbb{E}U(\tilde{\pi}, R, \lambda), \quad (16)$$

where $\tilde{\pi} = (\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_N)$ satisfies following expected utility maximization,

$$\tilde{\pi} = \arg \max_{\pi} \mathbb{E}U(\pi, \tilde{R}, \lambda) \quad (17)$$

subject to

$$\pi' \mathbf{1} = 1, \quad \text{and} \quad \pi \geq 0,$$

where \tilde{R} is a random sample of size T drawn from the empirical distribution $\hat{F}(R)$. As we discussed in Section 2.3, estimation imprecision of the sample moments can be measured directly by resampling methods. Solving the optimization problem (17) using B sets of samples leads to B portfolios, $\tilde{\pi}_b$, $b = 1, 2, \dots, B$. The investor's strength of belief parameter τ can also be related to the degree of shrinkage and be expressed as, say the r -th quantile of the distribution of ξ , $0 < r < 1$, i.e.,

$$\tau = G^{-1}(r) \equiv \xi_r, \quad \text{say,}$$

where $G(\cdot)$ is distribution function of ξ . Thus, the first inequality constraint, $\mathbb{E}U(\pi, R, \lambda) \geq \tau$ in (15), can be represented as a confidence interval, $I = [\xi_r, \xi^U]$, where ξ^U is the same as the maximized expected utility of MV efficient portfolio given λ if $\mathbb{E}U(\cdot)$ is the quadratic expected utility function. This is due to the fact that when there is no estimation error, the maximized expected utility evaluated at these exact moments dominates all values generated by $\tilde{\pi}_b$, $b = 1, 2, \dots, B$.

The confidence interval has a nice interpretation as a measure of uncertainty (see

Bewley, 1988). Suppose an investor has high uncertainty aversion in the portfolio selection problem. Then, s/he will select relatively low τ , i.e., ξ_r with a small value of r , and use a $(1 - r)\%$ confidence interval. Since $\tau \equiv \xi_r$ represents an investor's strength of belief, we can correspond ξ_r with a large value of r , with investor who has less uncertainty in estimation, and vice-versa. Garlappi, Uppal, and Wang (2004) used the notion of the confidence interval to explain investor's aversion toward uncertainty using a multi-prior approach, and showed that their estimated portfolio weights shrink toward the weights of minimum variance portfolio more than those of empirical Bayes-Stein portfolio. While recent studies based on the empirical Bayes-Stein estimator tried to estimate admissible moments at the first stage and then optimize the portfolio weights by the MV principle, weights achieved by minimizing CE objective function subject to sets of constraints are shrunk directly to an appropriate prior weights, \mathbf{q} . Moreover, as Frost and Savarino (1986) emphasized, there is no certain way to select a particular informative prior in Bayesian decision rules. One can readily choose alternative informative priors for the Bayes-Stein estimator and obtain different type of shrinkage estimators for portfolio weights by calculating somewhat complex predictive density. However, instead of choosing alternative informative priors, one can choose an appropriate prior weight vector \mathbf{q} , and minimize the CE measure to estimate portfolio weights which also has the shrinkage interpretation. Thus, we can say that CE measure works directly as shrinkage estimator of portfolio weights in asset allocation problem.

The MV criterion has good performance as far as returns are driven by an elliptical distribution, such as, normal, Student's t and Levy distributions. Chamberlain (1983) showed that the MV approximation of the expected utility is exact for all utility functions for an elliptical distribution. Thus, for simplicity, we consider the maximization of the quadratic expected utility function given in (4), i.e.,

$$\max_{\pi} \mathbb{E}U(\pi, R, \lambda) = \max_{\pi} \left[\pi' m - \frac{\lambda}{2} \pi' \Sigma \pi \right] \quad (18)$$

subject to

$$\pi \geq 0, \quad \text{and} \quad \pi' \mathbf{1}_N = 1.$$

One can use bootstrap or Monte-Carlo methods to estimate a distribution of ξ in (16), i.e., resampling $T \times N$ samples for B times from the empirical distribution, $\hat{F}(R)$. Let these resampled series be $\tilde{R}_{(b)}$, $b = 1, 2, \dots, B$. Then, $\tilde{\pi}_{(b)}$ and $\tilde{\xi}_{(b)}$ can be calculated as follows

$$\tilde{\pi}_{(b)} = \arg \max_{\pi} \left[\pi' \tilde{m}_{(b)} - \frac{\lambda}{2} \pi' \tilde{\Sigma}_{(b)} \pi \right], \quad (19)$$

$$\xi_{(b)} = \tilde{\pi}'_{(b)} \hat{m} - \frac{\lambda}{2} \tilde{\pi}'_{(b)} \hat{\Sigma} \tilde{\pi}_{(b)}, \quad (20)$$

where \hat{m} and $\hat{\Sigma}$ are the sample mean and sample covariance matrix estimated from original return data R , and $\tilde{m}_{(b)}$ and $\tilde{\Sigma}_{(b)}$ are calculated from simulated data $\tilde{R}_{(b)}$. The empirical distribution of ξ can be estimated based on $\xi_{(b)}$, $b = 1, 2, \dots, B$. Then, the CE minimization problem can be written as

$$\min_{\pi} \sum_{i=1}^N \pi_i \ln(\pi_i/q_i) \quad (21)$$

subject to

$$\pi' \hat{m} - \frac{\lambda}{2} \pi' \hat{\Sigma} \pi \geq \hat{G}^{-1}(r), \quad \pi \geq 0, \quad \text{and} \quad \pi' \mathbf{1}_N = 1, \quad (22)$$

where $\hat{G}(\cdot)$ denotes the empirical distribution function of ξ . Under the assumption of smooth expected utility function in (18), it is straightforward to solve the optimization problem minimize (21)-(22) by classical gradient based routine. This is in contrast to the Leontief utility function discussed in Section 3.2, for which no easy solution is available.

Using the monthly data given in Michaud (1998, p.14) on eight international equities, Figure 2 shows the shrinkage effect of minimizing CE portfolio weights when \mathbf{q} is chosen as equally weighted portfolio. Points A , B , and C denote MV efficient, minimum CE and equally weighted portfolios, respectively, with $\lambda = 0.06$. Standard deviations and means (monthly) associated with these portfolios are (2.599, 1.131), (3.006, 1.146), and (3.459, 1.168), respectively. Solid line denotes maximized expected utility indifference curve under MV efficient portfolio, and broken line represents that of CE portfolio at 0.2 quantile level ($r = 0.2$). Mixed line is the MV efficient frontier. Statistically equivalent points, $(\sqrt{\tilde{\pi}'_{(i)} \hat{\Sigma} \tilde{\pi}_{(i)}} \tilde{\pi}'_{(i)} \hat{m})$, $i = 1, 2, \dots, 500$ for the MV efficient portfolio are represented by small dots. We note that minimization of CE shrinks MV efficient portfolio (point A) toward the more diverse equally weighted portfolio (point C). The degree of shrinkage depends on τ , the investor's degree of uncertainty aversion.

[Figure 2]

Figure 3 shows non-parametric kernel density for ξ based on 500 data points. The shape of the density is clearly negatively skewed. Since $\xi^{MV} = 1.131 - 0.06/2 \times 2.599^2 = 0.928$ and $\xi_{0.2} = 0.874$, 80% confidence interval is given by [0.874, 0.928].

It can be checked that $\xi_{0.2}^{CE} = 1.146 - 0.06/2 \times 3.006^2 \simeq 0.876$, is very close to the 0.2 quantile of ξ from Figure 3. That is, the maximized utility value of point B in Figure 2 is “almost” the same as 0.2 quantile of ξ .

[Figure 3]

Next we use data from Kenneth French’s website (<http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/>). These are monthly equal-weighted returns for health, utility, and others industry portfolios for the period January 1970 to May 2005. For this data the mean and covariance matrix are given by $(1.551, 1.156, 1.215)'$ and

$$\begin{pmatrix} 57.298 & 12.221 & 33.026 \\ 12.221 & 13.168 & 11.814 \\ 33.026 & 11.814 & 27.952 \end{pmatrix},$$

respectively. Figure 4 shows contour curves of $-\sum_{i=1}^N \pi_i \ln \pi_i$ for the three assets ($N = 3$) on the monthly portfolio standard deviation-mean plane. We consider every possible combination of weights π_1 , π_2 and π_3 , each taking 50 equally spaced values in $(0, 1)$, and satisfying $\sum_{i=1}^N \pi_i = 1$. The upper envelope curve in Figure 4 corresponds to the set of MV efficient portfolio. The point where $\sum_{i=1}^N \pi_i \ln \pi_i$ takes highest value represents equally weighted portfolio. The smoothness of each contour curve ensures existence of a unique solution if we are to solve the minimization problem (21)-(22) with $\mathbf{q} = N^{-1}\mathbf{1}$. Figure 5 shows contour curves of $-\sum_{i=1}^N \pi_i \ln(\pi_i/\pi_i^{min})$, where π^{min} is portfolio weights for minimum variance portfolio. We can see that the largest value of the function corresponds to minimum variance portfolio. Since minimum variance portfolio does not take account of the portfolio mean value, contour graph of Figure 5 is sensitive to the mean values compared to that in Figure 4. Thus, by minimizing CE with $\mathbf{q} = \pi^{min}$, it shrinks toward minimum variance portfolio and at the same time takes care of the portfolio mean values.

[Figure 4]

[Figure 5]

4 Generalized cross entropy method

When asset managers are allowed to sell short, the models presented in the previous section cannot be used directly. Eliminating the no-short-selling constraints $\pi \geq 0$ from (15) might lead to non-existence of the objective function (14), since the function $\ln(\cdot)$ is defined only for non-negative values. In this situation generalized

cross entropy (GCE) method proposed by Golan, Judge and Miller (1996, p.77) can be used allowing for negative portfolio weights for some assets. Let us define a discrete probability distribution $p_i = (p_{i1}, p_{i2}, \dots, p_{iM})'$, $i = 1, 2, \dots, N$ over $[l, u]$, a set of equally distanced discrete points $z = (z_1, z_2, \dots, z_M)'$. Similarly, let $\omega_i = (\omega_{i1}, \omega_{i2}, \dots, \omega_{iM})'$ be a discrete prior probability distribution for each prior q_i over z . The portfolio weights, then, can be represented by

$$\pi = \mathbf{Z}p = \begin{bmatrix} z' & 0 & 0 & 0 & 0 \\ 0 & z' & 0 & 0 & 0 \\ 0 & 0 & z' & 0 & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & 0 & z' \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ \vdots \\ p_N \end{bmatrix}.$$

In the quadratic expected utility case, we can consider the following GCE minimization problem instead of the CE minimization problem given in (21)-(22) to allow for short-selling, i.e.,

$$\min_{p \in \mathcal{P}} \sum_{i=1}^N \sum_{m=1}^M p_{im} \ln(p_{im}/\omega_{im}) \quad (23)$$

subject to

$$\begin{aligned} (\mathbf{Z}p)' \hat{m} - \frac{\lambda}{2} (\mathbf{Z}p)' \hat{\Sigma} (\mathbf{Z}p) &\geq \hat{G}^{-1}(r) \\ p_i' \mathbf{1}_M &= 1, \quad i = 1, 2, \dots, N, \\ (\mathbf{Z}p)' \mathbf{1}_N &= 1, \end{aligned}$$

where $\hat{G}(\cdot)$ is the empirical distribution function of the maximized expected utility ξ in (16). If we let \hat{p} be the solution to (23), each $\hat{\pi}_i$ for i -th asset ($i = 1, 2, \dots, N$) can be calculated by

$$\hat{\pi}_i = z' \hat{p}_i = \sum_{m=1}^M z_m \hat{p}_{im}.$$

At this stage, it is worthwhile to mention two important points about GCE portfolio selection problem. First, one has to set the support $[l, u]$ in such a way that the solution of GCE portfolio selection problem yields appropriate negative weights. For example, one can simply set $[l, u] = [-1, 1]$ and consider 11 equally distanced discrete points, $z = (-1.0, -0.8, -0.6, \dots, 0.6, 0.8, 1.0)'$. However, it may not lead to appropriate weights if the support $[l, u]$ is not wide enough to generate the MV portfolio weights. Since GCE portfolio is equivalent to that of MV efficient portfolio when input estimates are exact, the MV efficient portfolio should be in the set $\{\pi | \pi = \mathbf{Z}p, p \in \mathcal{P}\}$. Note that, theoretically, MV portfolio weights can be any numbers in

the real line. Thus, the support $[l, u]$ has to be wide enough to generate MV portfolio weights. Second, ω_{im} in (23) can be interpreted as the given discrete prior for the original prior q_i over z , which should be determined before estimation procedure. However, only one q_i for each $i = 1, 2, \dots, N$ is known before estimation stage, for example, it can be that of minimum variance or equally weighted portfolio. Thus, choosing ω_{im} , $m = 1, 2, \dots, M$ for a q_i is not easy. We choose $\omega_i = (\omega_{i1}, \omega_{i2}, \dots, \omega_{iM})'$ using the ME principle, and maximize $-\sum_{m=1}^M \omega_{im} \ln \omega_{im}$ with respect to ω_i and side conditions $\sum_{m=1}^M z_m \omega_{im} = q_i$ and $\sum_{m=1}^M \omega_{im} = 1$. This ω_i provides a most uniform (largest variance) probability distribution and an uncertainty measure for each of the q_i ($i = 1, 2, \dots, N$) over z . Therefore, our choice is not so arbitrary, and as we will see in our empirical application, it works quite well. The unique solution of the above optimization problem will have expressions similar to those in (11).

5 Empirical application

To illustrate the practical usefulness of our methodology, we consider an application of maximum entropy portfolio selection approach using eight international equity indices. The returns are computed from the month-end US dollar value for the period, December 1969 to July 2005. The indices are for the United States, Canada, Italy, Japan, United kingdom, Switzerland and Germany. Data are from Morgan Stanley Capital International (MSCI). The number of observation is 428. Summary statistics for the data are presented in Table 1.

[Table 1]

We compare the performance of the following asset allocation models discussed earlier: MV efficient portfolio (Markowitz, 1952) (MV); empirical Bayes portfolio (Jorion, 1986) (EB); Bayes with diffuse prior (BDP); minimum variance portfolio (MinV); equally weighted portfolio (EQ); resampled efficient portfolio (Michaud, 1998) (RS); two cases of cross entropy (CE) portfolio, one (CE_1) with prior weight vector \mathbf{q} corresponding to the equally weighted portfolio and for the other (CE_2), \mathbf{q} comes from the minimum variance portfolio. For all models except for EQ both with- and without-short-sale case are considered. Those with-short-sales, the portfolios are computed using generalized CE (GCE), and will be denoted by MVs, EBs, BDPs, MinVs, RSs, CE_{s1} and CE_{s2} , respectively. Since Frost and Savarino (1986) and Jorion (1986) used the empirical Bayes procedure with shrinkage toward equally weighted and minimum variance portfolios, respectively, we expect CE_1 and CE_2 to generate similar results to theirs.

In order to analyze portfolio performance we use “rolling window” scheme. We consider four window lengths, $W = 24, 48, 60, 120$, months and estimate parameter values over each W and all asset allocation models. Using the estimated parameters, optimal portfolio for each considered model is calculated. The portfolio return for next period can then be obtained by holding the portfolio with these weights over the next period. Since we deal with monthly return data out-of-sample holding period is a month. We repeat this procedure by moving the window for the next period, i.e., dropping the observation for the beginning month and including the data for the next month until we reach the last (428th) observation.

To evaluate the performance of each model, we use two evaluation measures, the Sharpe ratio (SR) and the certainty equivalent return (CEQ). Each evaluation measure is calculated at both in- and out-of-sample cases. For the in-sample case, evaluation measures are based on the estimated parameters over the chosen window. The average of in-sample estimate of the SR is given by

$$SR_{in} = \frac{1}{(T - W)} \sum_{t=W}^T \frac{\hat{\pi}'_t \hat{m}_t}{\sqrt{\hat{\pi}'_t \hat{\Sigma}_t \hat{\pi}_t}}, \quad (24)$$

where \hat{m}_t , $\hat{\Sigma}_t$, and $\hat{\pi}_t$ denote, respectively, the estimates of the mean and the covariance matrix, and the portfolio weight vector for the window, $[t - W + 1, t]$. For the out-of-sample case the returns of the resulting portfolio depend on the next period returns of each asset. Following rolling window scheme, the out-of-sample portfolio return at time $t + 1$ can be calculated by $\hat{\mu}_{t+1} = \hat{\pi}'_t R_{t+1}$, where R_{t+1} denotes the returns at time $t + 1$. The out-of-sample mean, variance, and SR of returns can be written, respectively, as

$$\tilde{m} = \frac{1}{(T - W)} \sum_{t=W}^T \hat{\mu}_t, \quad (25)$$

$$\tilde{\sigma}^2 = \frac{1}{(T - W - 1)} \sum_{t=W}^T (\hat{\mu}_t - \tilde{m})^2, \quad (26)$$

$$SR_{out} = \frac{\tilde{m}}{\tilde{\sigma}}. \quad (27)$$

For the other evaluation measure, the certainty equivalent return, we assume that the first and second moments of return can summarize an investor’s preference, and we define CEQ as

$$CEQ = \pi' m - \frac{\lambda}{2} \pi' \Sigma \pi, \quad (28)$$

where λ is the risk aversion parameter. The CEQ averages for the in- and out-of-sample cases are given, respectively, by

$$CEQ_{in} = \frac{1}{(T - W)} \sum_{t=W}^T \left(\hat{\pi}'_t \hat{m}_t - \frac{\lambda}{2} \hat{\pi}'_t \hat{\Sigma}_t \hat{\pi}_t \right), \quad (29)$$

and

$$CEQ_{out} = \tilde{m} - \frac{\lambda}{2} (\tilde{\sigma})^2, \quad (30)$$

where \tilde{m} and $\tilde{\sigma}$ are defined in (25) and (26).

We evaluate portfolio performance for five different values of the risk aversion parameter, namely, $\lambda = 0.07, 0.10, 0.17, 0.51, 1$. However, since the qualitative results regarding the comparison of different portfolio formation techniques are quite similar for all the values of λ , we present the results only for $\lambda = 0.10$. The results for other values of λ are available from authors. In Table 2, we present the results for window length $W = 24, 48$ and the results for $W = 60, 120$ are given in Table 3. For each window length there are some interesting common results: (i) when short-sales are not allowed, MV performs the best in terms of both SR and CEQ among all considered models for the in-sample case. When short-sales are allowed, MVs performs better than MV; their out-of sample performances, however, are very poor. We observe that SR and CEQ of MVs are uniformly lower than any other models for the out-of-sample case; (ii) EQ has higher values of out-of-sample SR and CEQ than MVs. This implies that classical MV portfolio's out-of-sample performances are not good. These results agree with those of Jorion (1985) and DeMiguel, Garlappi, and Uppal (2005) who compared the performances of EQ and MVs; (iii) as Frost and Savarino (1988) and Jagannathan and Ma (2003) demonstrated, imposing short-sales constraints helps to improve the out-of-sample performance for MV, EB and MinV; (iv) for CE_1 and CE_2 , as expected, the in-sample SR and CEQ values monotonically increase with the value of r , i.e., as the degree of investor's belief for the sample mean and covariance increases. In-sample SR and CEQ of CE_2 are always higher than those of MinV, and also those of EB for certain high values of r . This is due to the fact that the degree of shrinkage effects of CE_2 at certain high quantile values is lower than those of EB. For example, since CE_2 works as shrinkage rule from MV to MinV, resulting values of SR and CEQ should be located between those of MV and MinV. The same argument applies to CE_1 . SR and CEQ values of CE_1 should be between those of MV and EQ. We can see that for the in-sample case, SR and CEQ values move toward those of MV as r increases. And as r decreases, SR and CEQ values of CE_1 and CE_2 move toward those of EB and EQ, respectively. However, for the out-of-sample situations, we do not notice any particular orderings.

As we can see in Tables 1 and 2 the results for CE_{s_1} and CE_{s_2} are very similar to those of CE_1 and CE_2 . We, therefore, summarize the results for without-short-sale only. For small window length (W), mean and covariance estimates are likely to have large estimation errors. When we have $W = 24$, the ratio of W to N , $W/N = 24/8 = 3$, is relatively low, and this case could be thought of as asset allocation problems with relatively large number of assets (say, $N = 500$ and $W = 1500$). As Table 2 shows, EQ has higher out-of-sample SR and CEQ values than those of MV, and, moreover, SR values of EQ is even higher than those of EB and MinV. These surprising results are due to imprecision of sample mean and sample covariance. Indeed, if one assume all the assets have the same mean, variance and correlation, the resulting optimal portfolio is EQ portfolio. In such a case, the out-of-sample performance can be improved by choosing \mathbf{q} that of EQ rather than of MinV portfolio.

The out-of-sample CEQ of CE_1 for $r = 0.5$, $W = 24$ is 0.0722 which is the second highest value among all considered models. On the other hand, the poor out-of-sample performance of CE_2 shows that choosing MinV portfolio as \mathbf{q} is not enough to improve the performance. As we increase the window length W from 24 to 48, 60 and 120, we find that CEQ values of CE_1 are lower than that of MV (Tables 2 and 3). This better performance of MV is due to increased accuracy of the sample covariance estimates with relatively larger number of observations. For larger value of W , the performance of CE_2 is also much improved due to lower sampling errors and shrinkage towards the MinV portfolio.

When $W = 60$, $CE_2(r = 0.5)$ has the highest out-of-sample SR and CEQ, and the difference of CEQ values between $CE_2(r = 0.5)$ and EB is $0.3193 - 0.3024 = 0.0169$. Also all SR and CEQ values of CE_2 are higher than those of CE_1 . When $W = 120$, $CE_2(r = 0.2)$ performs the best. The difference of CEQ values between $CE_2(r = 0.2)$ and EB is $0.3197 - 0.3145 = 0.0052$, which is lower than 0.0169 (for $W = 60$). This decrease may be due to reduced in sampling errors resulting from larger window length.

Michaud (1998)'s resampled efficient portfolio (RS) performs relatively well when W is small, however, as W increases, the performance becomes worse. Since RS is calculated by taking sample average of resampled portfolios, it leads to well-diversified allocation, and it shares similar diversification characteristics of CE_1 . However, since with larger sample size, the sample covariance can be estimated with high degree of precision more diversified portfolios may not lead to improved CEQ values. From Table 3, for $W = 120$, the CEQ of RS is 0.2364 which is better than the CEQ of CE_1 for all values of r but lower than those of CE_2 . On the other hand, the out-of-sample performance of RSs are not as good as RS and very similar to MVs for all values of W .

Overall, we can say that our CE_1 portfolios perform better in terms of SR and CEQ values than the classical MV and EB procedure with small number of observations. With relatively large number of observations, we can estimate the covariance matrix with more precision, and in that case CE_2 portfolios perform very well.

[Table 2]

[Table 3]

To get an idea of the structure of portfolios obtained using our CE minimization technique, we plot the weights that would be assigned to the United States market over the each out-of-sample period. Four components of Figure 6 represent portfolio weights of the United States market in case of no short-selling. In Figure 6(i) we present weights for $CE_1(r = 0.2, 0.5)$ and MV models when $W = 24$. We note that although the direction of fluctuations of weights are very similar, both the weights of CE_1 are more stable than those of MV which have high fluctuation over the whole interval $[0, 1]$. Weights of CE_1 vary roughly above $1/8 = 0.125$ (the equal weight with 8 assets), shown by a solid horizontal line in the graph. This leads to higher SR and CEQ values of CE_1 . $CE_1(r = 0.5)$ weights are relatively more volatile than for $r = 0.2$. The later case represents investor's higher degree of uncertainty that leads to more shrinkage toward equally weighted portfolio. Other graphs in Figure 6 are self-explanatory. Briefly, $CE_1(r = 0.5)$ weights are more stable than EB weights and are closer to those of RS in (ii). In Figure 6(iii), where we display the graph for $W = 120$, we note that with larger window length, MV weights are relatively less volatile than what we noted in (i) and (ii), compared to those of CE_2 for $r = 0.2$ and 0.1 , both of which give almost identical result. Finally, Figure 6(iv) shows that the weights of EB are almost identical to those of $CE_2(r = 0.2)$ as expected, however, EB has smaller SR and CEQ values than $CE_2(r = 0.2)$ as we noted earlier from Table 3.

In Figure 7, we report generalized cross entropy (GCE) portfolio weights of the United States market when short-selling is allowed, i.e., without putting positivity constraints to the portfolio weights. From Figure 7(i) and (ii) with $W=24$, most of CEs_1 have roughly the same weight as in the equally weighted portfolio, i.e., 0.125. However, for the MVs, RSs and EBs, the degree of fluctuation of portfolio weights tends to increase with short-selling. For $W = 120$ in Figures 7(iii) and (iv), the weights for GCE portfolios are more stable compared to those of MVs and RSs. This behavior is quite different from those obtained in Figures 6(iii) and (iv) with no-short-sale.

[Figure 6]

[Figure 7]

6 Concluding Remarks

The Markowitz MV portfolio optimization theory is based on exact values of means, variances and covariances of assets. When the sample mean and covariance matrix are used to calculate portfolios in MV principle, the portfolio weights have extreme values and out-of-sample performances are not very good. To take care of these shortcomings, many empirical Bayes-Stein type estimation approaches have been proposed in the literature. These are known as shrinkage estimation, and they perform relatively better. However, there are many ways to choose the prior. Depending on chosen prior, resulting predictive densities will be different. And also derivations and estimations of predictive densities sometimes require complex procedures. We provide an alternative way of portfolio selection model by introducing cross-entropy (CE) and generalized CE (GCE) as the objective functions. Since CE and GCE measures can be also interpreted as shrinkage rule, our methods can be thought of as *direct shrinkage* towards any reasonable portfolio. The degree of shrinkage is given by certain quantile values of resampled maximized (quadratic) expected utility which is designed to capture the imprecision of the mean and covariance matrix estimates.

Our empirical results demonstrate that the out-of-sample performances of our suggested portfolio selection procedure, given certain quantile values of maximized expected quadratic utilities, are superior to those of the classical MV or empirical Bayes investment rules.

There are two notable aspects of our proposed portfolio selection procedure. First, the prior (target) portfolio weights can be chosen freely. One can choose more reasonable prior weights whose efficiency is investment relevant. For example, a capitalization-weighted prior might be a good candidate in practice. Second, our method can be immediately extended to the more general utility function that incorporates higher moments, such as, asymmetry and leptokurtosis of asset returns. And that we would like to pursue in our future research.

References

- Bawa, V.S., Brown, S., Klein, R. (1979). *Estimation Risk and Optimal Portfolio Choice*. North Holland, Amsterdam.
- Bera, A.K., Biliias, Y. (2002). The MM, ME, EL, EF and GMM approaches to estimation: a synthesis. *J. Econometrics*. 107:51-86.
- Best, M.J., Grauer, R. (1991). On the sensitivity of mean-variance efficient portfolios to changes in asset means: some analytical and computational results. *Rev. Financ. Studies*. 4:315-342.
- Bewley, T. (1988). Knightian decision theory and econometric inference. *Working Paper*. Yale University.
- Chamberlain, G. (1983). A characterization of the distributions that imply mean-variance utility functions. *J. Econ. Theory*. 29:185-201.
- DeMiguel, V., Garlappi, L., Uppal, R. (2005). How inefficient is the 1/N asset-allocation strategy? *Working Paper*. London Business School.
- Fernholz, E. R. (2002). *Stochastic Portfolio Theory*. Springer-Verlag, New York.
- Frost, P.A., Savarino, J.E. (1986). An empirical Bayes approach to portfolio selection. *J. Financ. and Quanti. Analysis*. 21:293-305.
- Frost, P.A., Savarino, J.E. (1988). For better performance: constrain portfolio weights. *J. Portfolio Management*. 15:29-34.
- Garlappi, L., Uppal, R., Wang, T. (2004). Portfolio selection with parameter and model uncertainty: a multi-prior approach. *Working Paper*. University of British Columbia.
- Golan, A., Judge, G., Miller, D. (1996). *Maximum Entropy Econometrics Robust Estimation with Limited Data*, Wiley.
- Green, R., Hollifield, B. (1992). When will mean variance efficient portfolios be well-diversified. *J. Finance*. 47:1785-1809.
- Hoskisson, R.E., Hitt, M.A., Johnson, R.H., Moesel, D. (1993). Construct validity of an objective (entropy) categorical measure of diversification strategy. *Strategic Management J.* 14:215-235.
- Jagannathan, R., Ma, T. (2003). Risk reduction in large portfolio: why imposing the wrong constraints helps. *J. Finance*. 54:1651-1684.

- Jaynes, E.T. (1963). Information theory and statistical mechanics. In: *Statistical Physics*. New York, pp. 181-218.
- Jobson, J. D., Korkie, B. M. (1980). Estimation for Markowitz efficient portfolios. *J. Am. Statist. Assoc.* 75:544-554.
- Jorion, P. (1985). International portfolio diversification with estimation risk. *J. Business*. 58:259-278.
- Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *J. Financ. and Quanti. Analysis*. 21:279-292.
- Kullback, S., Leibler, R.A. (1951). On information and sufficiency. *Ann. Math. Statist.* 22:79-86.
- Ledoit, O., 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, O., Wolf, M. (2004a). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Analysis*. 88:365-411.
- Ledoit, O., Wolf, M. (2004b). Honey, I shrunk the sample covariance matrix. *J. Portfolio Management*. 30:110-119.
- Lubatkin, M., Merchant, H., Srinivasan, N. (1993). Construct validity of some unweighted productcount diversification measures. *Strategic Management J.* 14:433-449.
- Markowitz, H. (1952). Portfolio selection. *J. Finance*. 7:77-91.
- Maasoumi, E. (1993). A compendium to information theory in economics and econometrics. *Econometric Rev.* 12:137-181.
- Merton, R.C. (1972). An analytic derivation of the efficient portfolio frontier. *J. Quanti. Analysis*. 7:1851-1872.
- Michaud, R.O. (1989). The Markowitz optimization enigma: is optimized optimal? *Financ. Analysts J.* 45:31-42.
- Michaud, R.O. (1998). *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. Boston: Harvard Business School Press.
- Scherer, B. (2002). Portfolio resampling: review and critique. *Financ. Analysts J.* 58:98-109.

- Sharpe, W. (1963). A simplified model for portfolio analysis. *Management Science*. 9:277-293.
- Shannon, C.E. (1948). The mathematical theory of communication. *Bell System Technical J.* July-Oct. Reprinted In: *The mathematical theory of communication*. Urbana: University of Illinois, pp. 3-91.
- Ullah, A. (1996). Entropy, divergence and distance measures with econometric applications. *J. Statistical Planning and Inference*. 49:137-162.
- Zellner, A., Chetty, V.K. (1965). Prediction and decision problems in regression model from the Bayesian point of view. *J. Am. Statist. Associ.* 60:608-615.

TABLES

Table 1: Sample means, variances and the correlation matrix

Country	USA	Canada	Italy	France	Japan	UK	Swiss	Germany
Mean	0.9707	0.9823	0.7719	1.0347	1.1508	0.9789	1.1886	1.0062
Variance	19.9114	31.0974	53.0035	42.2422	40.6479	40.4327	28.5035	38.2843
Correlation	1.0000							
	0.7254	1.0000						
	0.2987	0.3479	1.0000					
	0.5025	0.4867	0.4949	1.0000				
	0.3104	0.3281	0.3485	0.3924	1.0000			
	0.5412	0.5239	0.3791	0.5748	0.3737	1.0000		
	0.5253	0.4786	0.4095	0.6390	0.4273	0.5793	1.0000	
	0.4762	0.4154	0.4601	0.6719	0.3634	0.4778	0.6871	1.0000

Table 2: In- and out-of-sample performance of asset allocation models

	$W = 24$				$W = 48$			
	In-Sample		Out-of-Sample		In-Sample		Out-of-Sample	
	SR	CEQ	SR	CEQ	SR	CEQ	SR	CEQ
	Short sales not allowed							
MV	0.3773	0.5253	0.2333	-0.0300	0.3416	0.4614	0.2359	0.0853
BDP	0.3768	0.5241	0.2339	-0.0190	0.3414	0.4612	0.2362	0.0883
EB	0.3149	0.2535	0.2257	0.0567	0.2872	0.2567	0.2420	0.1463
EQ	0.2201	-0.4728	0.2339	0.0513	0.2177	-0.2108	0.2358	0.0579
MinV	0.2556	0.0253	0.2211	0.0593	0.2676	0.1833	0.2413	0.1386
RS	0.3510	0.3879	0.2452	0.0782	0.3257	0.3914	0.2382	0.1062
$CE_1 (r = 0.1)$	0.2359	-0.3212	0.2340	0.0593	0.2410	-0.0602	0.2328	0.0554
$CE_1 (r = 0.2)$	0.2587	-0.1537	0.2362	0.0712	0.2620	0.0609	0.2314	0.0597
$CE_1 (r = 0.5)$	0.3101	0.1630	0.2379	0.0722	0.2981	0.2537	0.2321	0.0773
$CE_1 (r = 0.8)$	0.3466	0.3655	0.2392	0.0573	0.3220	0.3691	0.2369	0.1016
$CE_1 (r = 0.9)$	0.3583	0.4284	0.2396	0.0480	0.3300	0.4066	0.2394	0.1114
$CE_2 (r = 0.1)$	0.2676	0.0671	0.2162	0.0284	0.2689	0.1879	0.2416	0.1445
$CE_2 (r = 0.2)$	0.2752	0.0996	0.2102	-0.0076	0.2729	0.2028	0.2401	0.1425
$CE_2 (r = 0.5)$	0.3031	0.2208	0.2048	-0.0649	0.2930	0.2787	0.2392	0.1455
$CE_2 (r = 0.8)$	0.3373	0.3711	0.2102	-0.0718	0.3173	0.3702	0.2378	0.1271
$CE_2 (r = 0.9)$	0.3511	0.4286	0.2139	-0.0643	0.3261	0.4058	0.2394	0.1281
	With short sales							
MVs	0.5544	1.6812	0.1646	-2.1317	0.4335	0.9282	0.1615	-0.7192
BDPs	0.5534	1.6708	0.1671	-1.9446	0.4330	0.9270	0.1632	-0.6844
EBs	0.4306	0.8403	0.1934	-0.2324	0.3271	0.4455	0.2123	0.0106
MinVs	0.2853	0.2428	0.1920	-0.1172	0.2848	0.2879	0.2133	0.0044
RSs	0.5553	1.4152	0.1651	-2.0414	0.4351	0.9065	0.1621	-0.6994
$CE_{s_1} (r = 0.1)$	0.2201	-0.4728	0.2339	0.0513	0.2177	-0.2108	0.2358	0.0579
$CE_{s_1} (r = 0.2)$	0.2201	-0.4728	0.2339	0.0513	0.2194	-0.2035	0.2342	0.0528
$CE_{s_1} (r = 0.5)$	0.2202	-0.4707	0.2337	0.0508	0.2415	-0.0748	0.2230	0.0062
$CE_{s_1} (r = 0.8)$	0.2625	-0.1305	0.2234	0.0112	0.3014	0.2703	0.2135	-0.0227
$CE_{s_1} (r = 0.9)$	0.3306	0.2790	0.2174	-0.0469	0.3348	0.4425	0.2130	-0.0241
$CE_{s_2} (r = 0.1)$	0.2853	0.2428	0.1920	-0.1172	0.2848	0.2880	0.2133	0.0045
$CE_{s_2} (r = 0.2)$	0.2853	0.2428	0.1920	-0.1172	0.2848	0.2880	0.2133	0.0045
$CE_{s_2} (r = 0.5)$	0.2853	0.2428	0.1920	-0.1172	0.2889	0.3036	0.2101	-0.0082
$CE_{s_2} (r = 0.8)$	0.3078	0.3187	0.1856	-0.1589	0.3089	0.3914	0.1972	-0.0569
$CE_{s_2} (r = 0.9)$	0.3472	0.4964	0.1713	-0.2879	0.3306	0.4814	0.1910	-0.0933

Note: The table represents in-sample and out-of-sample results for $\lambda = 0.10$. SR and CEQ denote Sharpe ratio and certainty equivalence measure, respectively.

Table 3: In- and out-of-sample performance of asset allocation models

	$W = 60$				$W = 120$			
	In-Sample		Out-of-Sample		In-Sample		Out-of-Sample	
	SR	CEQ	SR	CEQ	SR	CEQ	SR	CEQ
	Short sales not allowed							
MV	0.3366	1.2129	0.2595	0.1961	0.3344	0.5165	0.2742	0.2432
BDP	0.3364	1.2089	0.2600	0.1998	0.3343	0.5165	0.2744	0.2447
EB	0.2890	0.8905	0.2763	0.3024	0.2955	0.3813	0.2819	0.3145
EQ	0.2244	0.7834	0.2625	0.1852	0.2470	0.0828	0.2514	0.1344
MinV	0.2755	0.8370	0.2768	0.3039	0.2891	0.3573	0.2815	0.3118
RS	0.3246	1.1538	0.2651	0.2279	0.3262	0.4830	0.2714	0.2364
$CE_1 (r = 0.1)$	0.2507	0.8729	0.2614	0.1898	0.2769	0.2450	0.2505	0.1409
$CE_1 (r = 0.2)$	0.2698	0.9376	0.2626	0.2033	0.2889	0.3053	0.2535	0.1583
$CE_1 (r = 0.5)$	0.3007	1.0462	0.2654	0.2254	0.3099	0.4079	0.2572	0.1788
$CE_1 (r = 0.8)$	0.3207	1.1253	0.2689	0.2458	0.3239	0.4720	0.2646	0.2099
$CE_1 (r = 0.9)$	0.3271	1.1550	0.2670	0.2366	0.3279	0.4898	0.2674	0.2210
$CE_2 (r = 0.1)$	0.2756	0.8374	0.2767	0.3035	0.2898	0.3598	0.2818	0.3134
$CE_2 (r = 0.2)$	0.2770	0.8434	0.2756	0.2995	0.2921	0.3683	0.2832	0.3197
$CE_2 (r = 0.5)$	0.2938	0.9167	0.2798	0.3193	0.3051	0.4165	0.2815	0.3111
$CE_2 (r = 0.8)$	0.3148	1.0258	0.2775	0.3048	0.3202	0.4715	0.2822	0.3071
$CE_2 (r = 0.9)$	0.3222	1.0759	0.2729	0.2800	0.3252	0.4890	0.2816	0.2994
	With short sales							
MVs	0.4086	1.7626	0.2076	-0.2586	0.3669	0.6468	0.2401	0.0243
BDPs	0.4083	1.7482	0.2091	-0.2400	0.3667	0.6467	0.2406	0.0289
EBs	0.3112	0.9385	0.2621	0.2425	0.3038	0.4218	0.2785	0.2985
MinVs	0.2862	0.8407	0.2662	0.2573	0.2962	0.3934	0.2798	0.3030
RSs	0.4103	0.7792	0.2113	-0.2498	0.3676	0.6441	0.2404	0.0247
$CEs_1 (r = 0.1)$	0.2244	-0.1300	0.2624	0.1849	0.2511	0.1017	0.2491	0.1247
$CEs_1 (r = 0.2)$	0.2266	-0.1208	0.2588	0.1696	0.2586	0.1422	0.2445	0.1052
$CEs_1 (r = 0.5)$	0.2517	0.0195	0.2457	0.1193	0.2873	0.2947	0.2385	0.0860
$CEs_1 (r = 0.8)$	0.3064	0.3135	0.2454	0.1349	0.3153	0.4307	0.2414	0.1059
$CEs_1 (r = 0.9)$	0.3314	0.4383	0.2423	0.1179	0.3274	0.4870	0.2426	0.1106
$CEs_2 (r = 0.1)$	0.2862	0.3062	0.2662	0.2572	0.2962	0.3934	0.2798	0.3030
$CEs_2 (r = 0.2)$	0.2862	0.3062	0.2662	0.2572	0.2967	0.3953	0.2792	0.3007
$CEs_2 (r = 0.5)$	0.2889	0.3162	0.2611	0.2359	0.3006	0.4099	0.2762	0.2898
$CEs_2 (r = 0.8)$	0.3027	0.3735	0.2465	0.1760	0.3123	0.4546	0.2677	0.2513
$CEs_2 (r = 0.9)$	0.3224	0.4523	0.2398	0.1400	0.3225	0.4932	0.2602	0.2145

Note: The table represents in-sample and out-of-sample results for $\lambda = 0.10$. SR and CEQ denote Sharpe ratio and certainty equivalence measure, respectively.

Figure 1: Maximum diversification problem with Leontief utility function

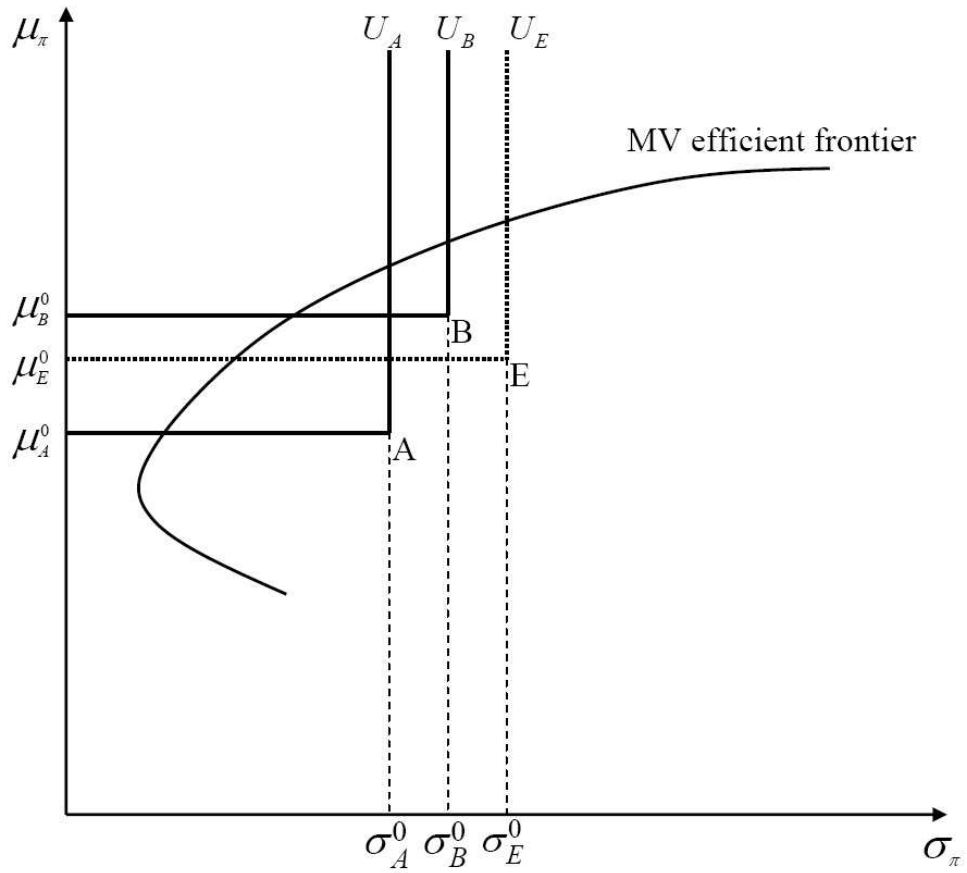
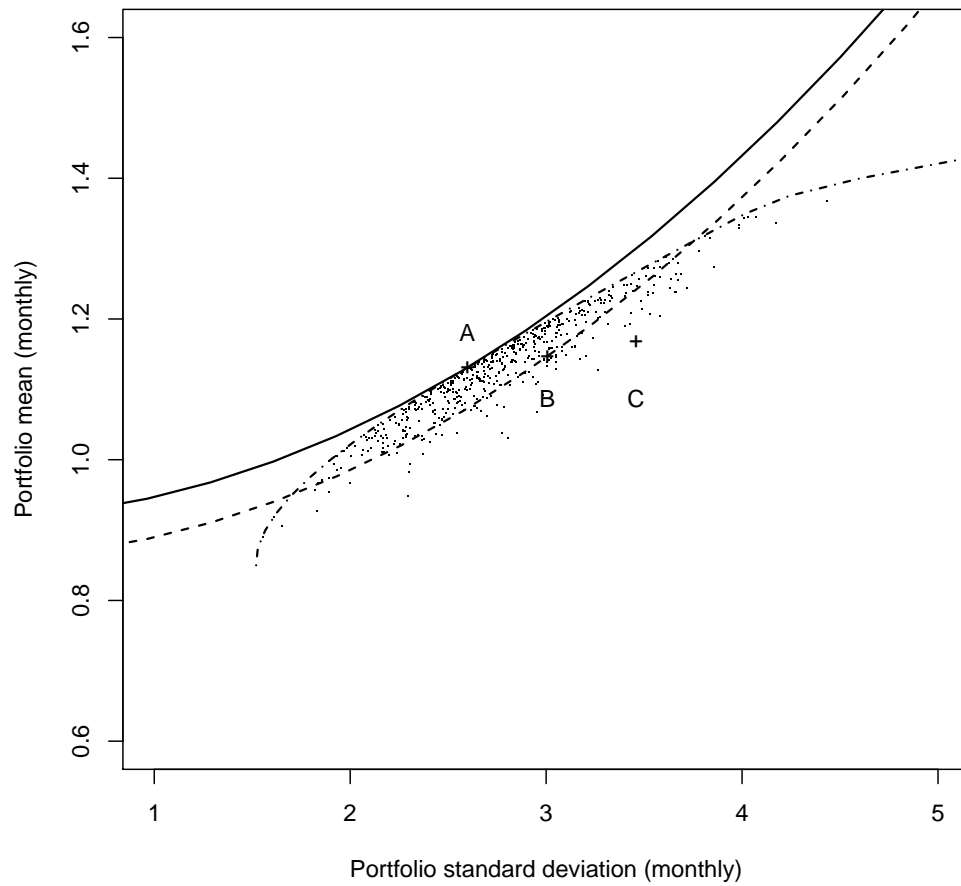
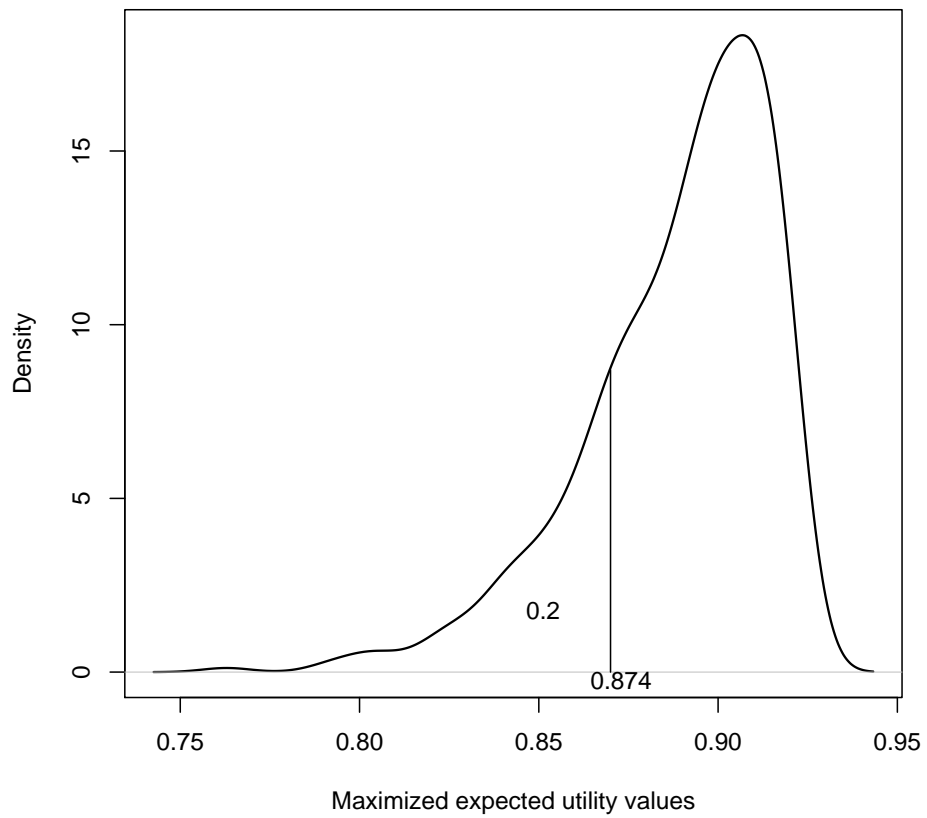


Figure 2: Shrinkage effects of minimizing cross-entropy



Note: The MV efficient frontier is illustrated by the mixed line. The 500 statistically equivalent portfolios associated with a MV efficient portfolio 'A' are represented by dot points. 'B' and 'C' denote minimum CE and equally weighted portfolios, respectively.

Figure 3: Non-parametric density for ξ



Note: Non-parametric kernel density for ξ is estimated based on 500 data points in Figure 2, and using optimal bandwidth = 0.0068.

Figure 4: Contour curves of $-\sum_{i=1}^N \pi_i \ln \pi_i$

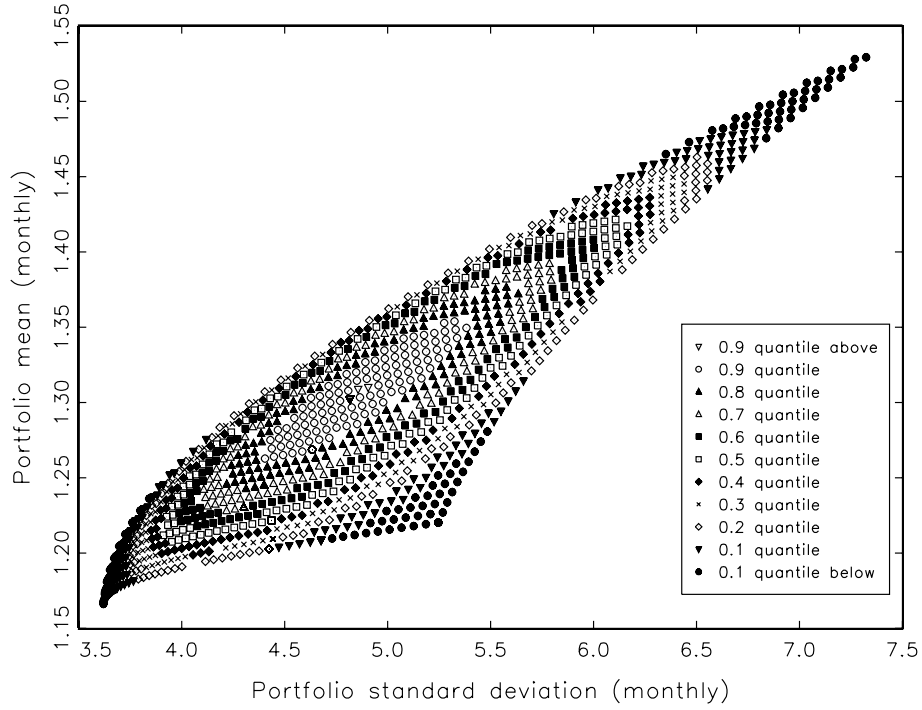


Figure 5: Contour curves of $-\sum_{i=1}^N \pi_i \ln(\pi_i/\pi_i^{min})$

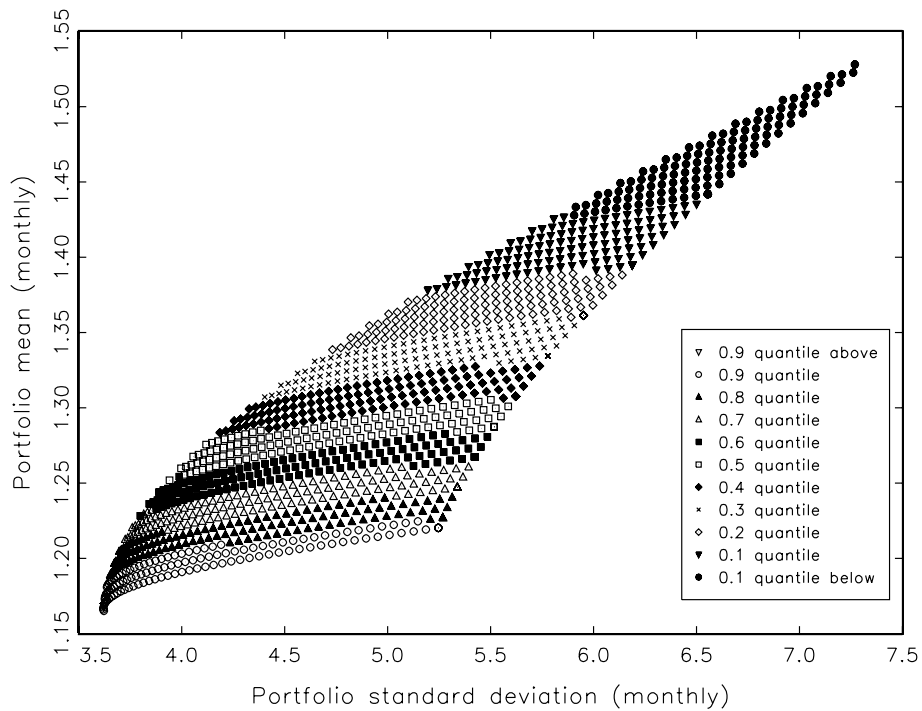
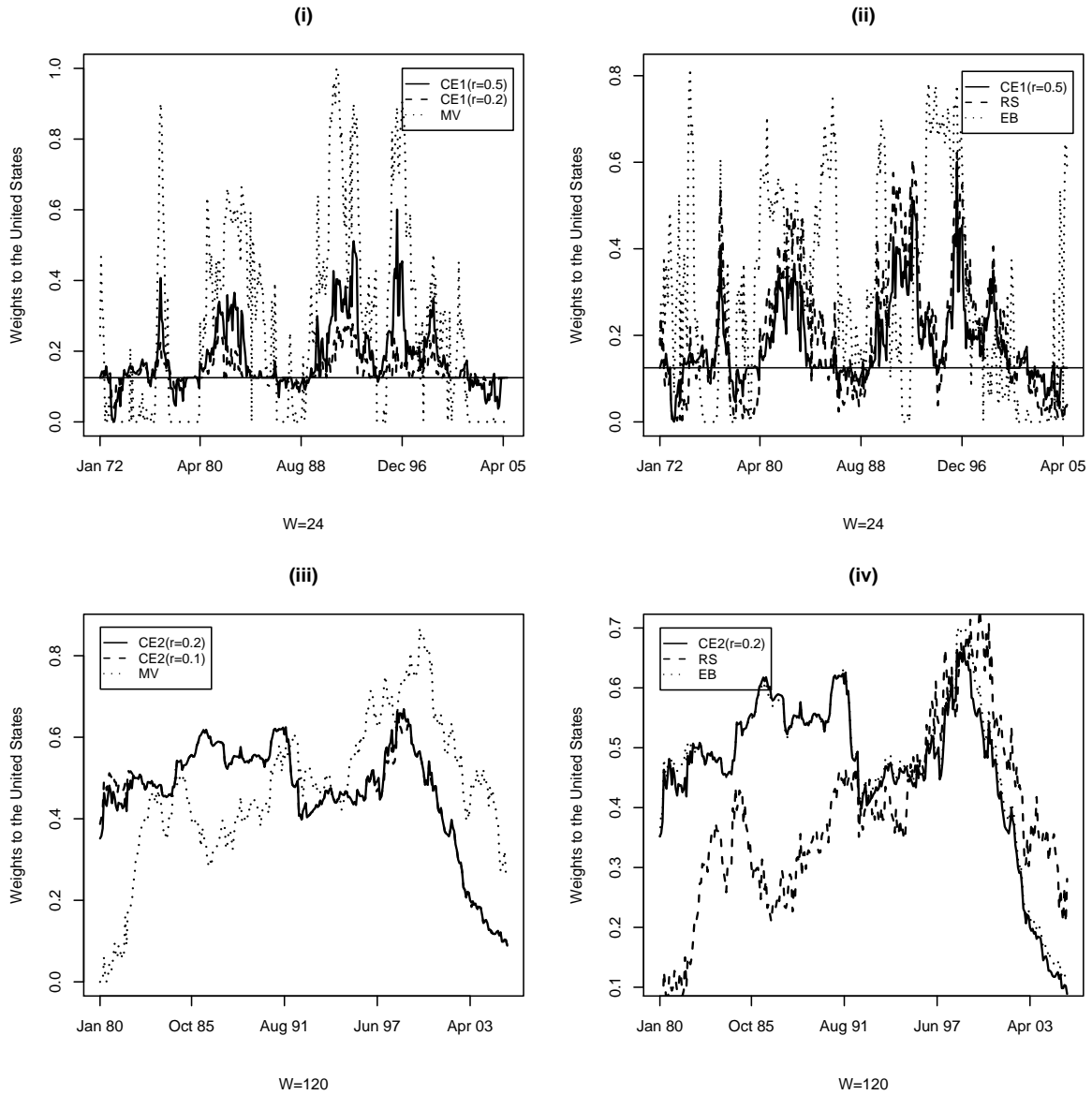
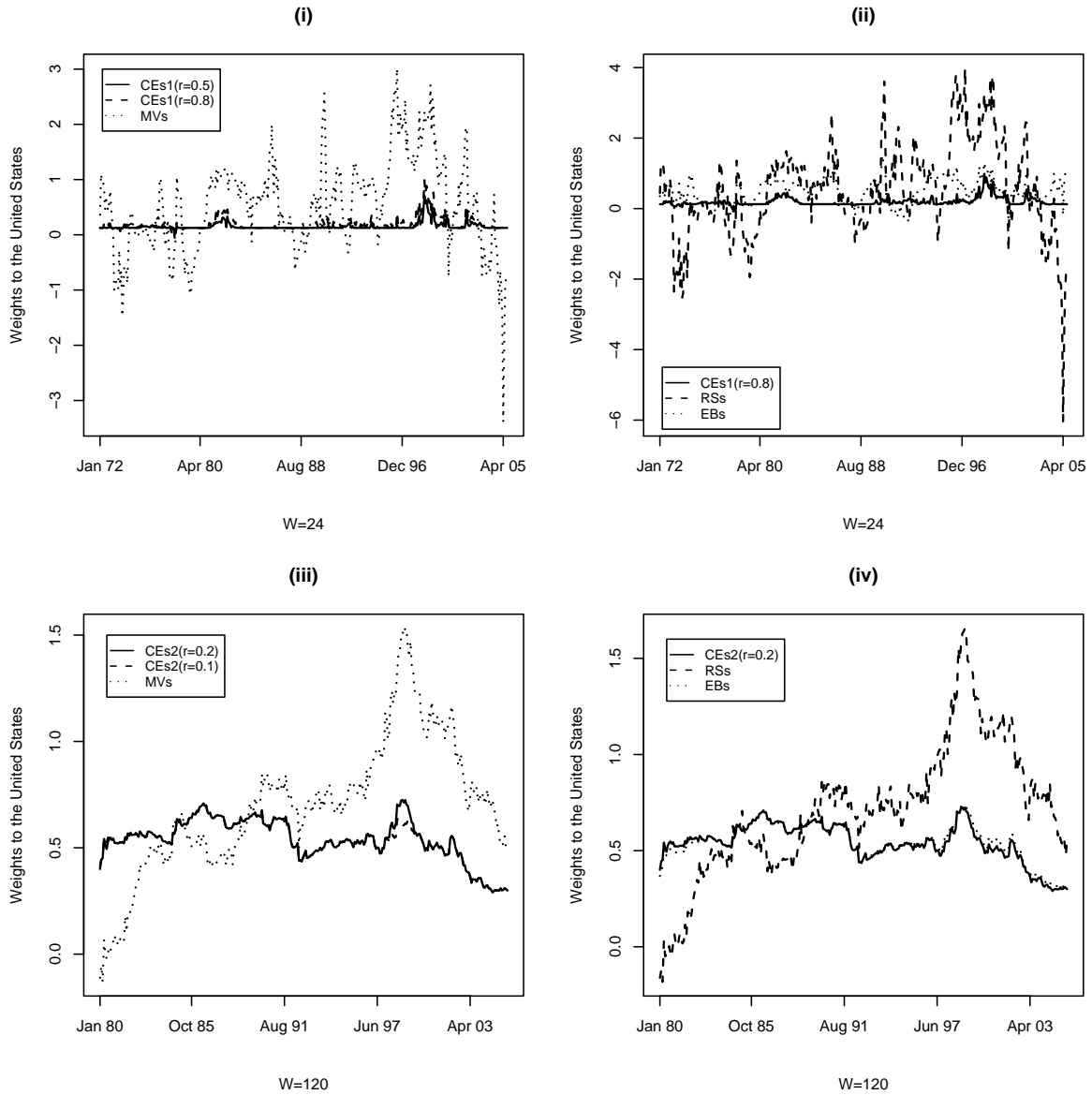


Figure 6: Weights to the United States: without short-selling



Figures(i)-(ii) and (iii)-(iv) illustrate portfolio weights assigned to the United States market for $W = 24$ and $W = 120$, respectively, when short-selling is not allowed.

Figure 7: Weights to the United States: with short-selling



Figures(i)-(ii) and (iii)-(iv) illustrate portfolio weights assigned to the United States market for $W = 24$ and $W = 120$, respectively, when short-selling is allowed.