



Munich Personal RePEc Archive

Bayesian Portfolio Selection with Gaussian Mixture Returns

Qian, Hang

Iowa State University

January 2009

Online at <https://mpra.ub.uni-muenchen.de/32688/>
MPRA Paper No. 32688, posted 09 Aug 2011 01:57 UTC

Bayesian Portfolio Selection with Gaussian Mixture Returns

Hang Qian

Iowa State University

January, 2009

Abstract

Markowitz portfolio selection is challenged by huge implementation barriers. This paper addresses the parameter uncertainty and deviation from normality in a Bayesian framework. The non-normal asset returns are modeled as finite Gaussian mixtures. Gibbs sampler is employed to obtain draws from the posterior predictive distribution of asset returns. Optimal portfolio weights are then constructed so as to maximize agents' expected utility. Simple experiment suggests that our Bayesian portfolio selection procedure performs exceedingly well.

1. Introduction

Markowitz (1952) mean-variance analysis laid out the foundation of portfolio selection theory. Optimal portfolio is constructed so as to minimize the portfolio variance for a targeted expected return, or in its dual form, to maximize expected return with controlled risk bearing. However, the classical framework is challenged by great implementation difficulties.

Firstly, the implementation is hampered by the inability to provide the inputs—the population mean and covariance matrix of asset returns are unknown. One might resort to the sample analog as an expedient, which is termed as “certainty equivalent solution”. It is well documented that under that approach, the portfolio weights tend to be quite sensitive to the minute change of inputs. The resulting portfolio features extreme volatility and lack of diversification (Dickinson ,1974; Jobson ,1980; Black and Litterman, 1992; Michaud,1998, among others). The instability of portfolio weights is, to a large extent, ascribed to the negligence of parameter uncertainty, especially estimation error of the mean. Chopra and

Ziembra (1993) find that error in means is ten-fold as devastating as errors in variances. Furthermore, assets with high sample average return and low sample variance obviously will be given larger weight in the portfolio. However, those assets returns are more likely to be error-ridden. (Scherer, 2002)

Secondly, to syncretize Markowitz mean-variance optimization with investors' expected utility maximization, it is usually assumed that the asset returns have a multivariate normal distribution. As long as we maintain some general form of concave utility functions other than quadratic, departing from normality is not innocuous. In that case, portfolio mean and variance may not fully characterize the portfolio distribution (even if the first two moments may be sufficient at the individual asset level, say Gamma distribution). However, the utility maximizing agents, in general, care about the entire distribution of the portfolio return, including higher moments such as skewness (Kraus and Litzenberger, 1976; Harvey and Siddique, 2000). It is possible that investors optimally choose a portfolio other than those on the mean-variance frontier. For example, for a given variance, an investor may further trade expected return with positive skewness, like buying a lottery.

The normality assumption is untenable from the empirical point of view. Investors usually feel the stock prices crawling upwards for months and plummet in a day. The magnitude of crash on Black Monday of 1987 is hard to reconcile with the relative thin tail of normal distribution. It is common for high-frequent financial data exhibiting asymmetry, Leptokurtosis and extreme value. Therefore, to better predict asset returns and select optimal portfolio it is worthwhile to adopt some more flexible functional form to characterize asset returns distribution.

One related point merits some remarks. That is, the mean-variance optimization *per se* does not depend on the normality assumption. Normality is required only when we want to justify the mean-variance argument to be consistent with utility maximization. For instance, suppose the goal of investment is simply to track a given portfolio—minimizing the variance of tracking error (possibly subject to attaining a minimum expected return). In that context, the investors care nothing but the first two moments of the portfolio. Furthermore, the only inputs required are mean and covariance matrix of individual asset returns vector—it does not matter whether it is normally distributed or not. However, here investors' utility comes from minimized tracking error variance, not necessarily from the maximization of end-of-period wealth. In this paper, we confine our analysis to the expected monetary utility maximizer, excluding those who minimize variance for the sake of hating variance. For those agents blessed with some general utility functions, the concern on normality is legitimate.

In this paper, we simultaneously address the estimation risk as well as non-normality returns in a unified Bayesian portfolio selection framework, in which estimation risk is absorbed

into the posterior distribution, and will manifest itself in the predictive returns distribution. On the other hand, non-normality is reflected in the finite Gaussian mixture likelihood function. Gibbs sampler is employed to obtain draws from the posterior, which can be further used to generate draws coming from the predictive returns distribution. By this stage, Bayesian mean-variance frontier is readily available. As we have argued, the mean-variance frontier could be problematic in the absence of normal returns. We therefore circumvent the frontier by directly working on expected utility maximization. With predictive returns draws in hand, we can simulate the expected utility and maximize it with respect to the portfolio weight.

2. Literature

Estimation risk in Markowitz portfolio selection can be tackled by a variety of strategies, such as robust optimization approach advocated by Goldfarb and Iyengar (2003), James-Stein estimator (e.g. Jorion, 1986), Portfolio resampling (e.g. Scherer, 2002), etc. Paralleled with frequentist solutions, Bayesians cope with the parameter uncertainty in a more natural manner. The prior information on asset returns, combined with historical data, yields the posterior as well as the predictive returns distribution, the mean and variance of which readily serve as the inputs for Markowitz portfolio selection. There is a substantial amount of literature using the Bayesian approach. Early theoretical work includes Klein and Bawa (1976), Bawa et al. (1979), Frost and Savarino (1986).

More recently, with the fast advancement of computation techniques, large scale Bayesian portfolio selection becomes possible. Polson and Tew (2000) use a three stage hierarchical model to derive the posterior predictive covariance matrix. Furthermore, they introduce some dynamics by employing a rolling estimation window to rebalance the portfolio weights in order to alleviate the problem of nonstationarity. In that paper, they also claim that their location-scale mixture model can accommodate nonnormal daily returns. However, nonnormality is one observable feature of historical data. Since their location and scale parameters in the likelihood function do not change with time (in each estimation period), in that case nonnormality cannot be observed in data.

Greyserman et al. (2006) is another hierarchical Bayesian portfolio selection paper using MCMC to generate draws from the posterior and predictive distribution. They work on both mean-variance frontier and direct utility maximization via numerical optimization. In comparison with classical and James-Stein model, they conclude that the hierarchical Bayesian model performs significantly better.

On top of estimation risk, nonnormality is another concern on the portfolio selection. There are abundant of evidence suggesting skewness and fat-tailness of asset returns (e.g. Fama,1965; Praetz, 1972; Peiro, 1994, to name a few). One of the tractable remedies to the nonnormality is mixture models, which, according to Geweke(2007), could date back to the work of Newcomb (1886). Finite Gaussian mixture models provide a flexible representation that could mimic virtually all shapes of density, including multimodality, skewness and Leptokurtosis, etc. There is some success of using mixture models to predict stock returns (see Kon, 1984; Weigand, 2000).

However, recent years witness some downturn of mixture models in Bayesian works, largely due to the controversy on the identification issues. Obviously, finite Gaussian mixtures suffer from the invariance of relabeling, i.e. the permutation of the parameter vector across regimes will not change the likelihood function. In that case, interpretation of posterior is difficult and Gibbs sampler exhibits unusual properties. Celeux et al. (2000) argue that virtually the entirety of MCMC samplers do not converge. Jasra et al. (2005) pessimistically believe that Gibbs sampler is not always appropriate for mixture model.

It is Geweke (2007) who bring the mixture model back to the stage. He insightfully points out: as long as the function of interest is invariant with respect to permutation, Gibbs sampler does reliably recover the posterior. As Frühwirth-Schnatter(2001) observes that if one adds a Metropolis-within-Gibbs step by proposing a random permutation, the proposal will be accepted for sure, because relabeling parameters leave the likelihood function unchanged, hence the invariance of posterior. Geweke (2007) further exploits the invariant nature by suggesting a conceptual permutation-augmented simulator—any of the copies suffices, including the one from the simple Gibbs sampler.

3. Methodology

3.1 The model

We begin with an investor's problem: choosing a desirable portfolio to maximize her expected utility derived from stochastic yield of one dollar investment.

Assume that the asset universe consists of N assets with returns vector \mathbf{R}_t in period t .

Living in current period T , the investor is planning a period ahead.

$$\max_{\omega} E[u(\omega' \mathbf{R}_{T+1}) | \mathbf{R}^h]$$

$$\text{s.t. } \omega' \mathbf{1} = 1$$

where,

$\mathbf{R}^h \equiv \{\mathbf{R}_t\}_{t=1}^T$ is the history of realized asset returns as of time T

\mathbf{R}_{T+1} is the next-period stochastic return

$u(\cdot)$ is some concave utility function

$E[\cdot | \mathbf{R}^h]$ is the expectation operator conditional on returns history

$\mathbf{1}$ is a vector of ones

The investor uses Bayes Rule to update her belief so the optimization problem can be formulated as:

$$\max_{\omega} \int u(\omega' \mathbf{R}_{T+1}) \cdot p(\mathbf{R}_{T+1} | \mathbf{R}^h) d\mathbf{R}_{T+1}$$

$$\text{s.t. } \omega' \mathbf{1} = 1$$

$$p(\mathbf{R}_{T+1} | \mathbf{R}^h) = \int p(\mathbf{R}_{T+1} | \theta, \mathbf{R}^h) \cdot p(\theta | \mathbf{R}^h) d\theta$$

$$p(\theta | \mathbf{R}^h) \propto p(\theta) \cdot p(\mathbf{R}^h | \theta)$$

where,

$p(\mathbf{R}_{T+1} | \mathbf{R}^h)$ is the posterior predictive distribution of future returns

$p(\mathbf{R}_{T+1} | \theta, \mathbf{R}^h)$ is the conditional predictive distribution, usually taking the same function form as the likelihood function

$p(\theta | \mathbf{R}^h)$ is the posterior, proportional to the prior $p(\theta)$ times likelihood function $p(\mathbf{R}^h | \theta)$

The deviation of normality is captured by a finite Gaussian mixture. Suppose in period t , there are S potential regimes that might generate \mathbf{R}_t . Of course, each of the models differs in mean and covariance. Then the likelihood function in period t can be written as:

$$p(\mathbf{R}_t | \theta) = \sum_{s=1}^S \pi_s \cdot \phi(\mathbf{R}_t; \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)$$

where,

π_s the probability that the period- t data is generated by model s

$\phi(\mathbf{R}_t; \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)$ is the p.d.f. of multi-normal distribution $N(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)$ evaluated at \mathbf{R}_t

$$\boldsymbol{\theta} \equiv \{\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \pi_s\}_{s=1}^S$$

To implement MCMC, it is convenient to work on the latent regime representation of the Gaussian mixtures. Let the period- t unobservable state be τ_t . Apparently, $\tau_t \in \{1, 2, \dots, S\}$, and occurs with probability $\boldsymbol{\pi} \equiv (\pi_1, \pi_2, \dots, \pi_S)$

In the case of $\tau_t = s$, we have $\mathbf{R}_t = \boldsymbol{\mu}_s + \boldsymbol{\varepsilon}_{t,s}$, $\boldsymbol{\varepsilon}_{t,s} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_s)$

With augmented latent state, the likelihood function in period t can be written as:

$$p(\mathbf{R}_t | \boldsymbol{\theta}, \tau_t) = \sum_{s=1}^S \phi(\mathbf{R}_t; \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s) \cdot I(\tau_t = s)$$

where $I(\cdot)$ is the indicator function that take the value of one if the expression inside is true, and zero otherwise.

Define $\boldsymbol{\tau} \equiv \{\tau_t\}_{t=1}^T$

$$p(\mathbf{R}^h | \boldsymbol{\theta}, \boldsymbol{\tau}) = \prod_{t=1}^T \sum_{s=1}^S \phi(\mathbf{R}_t; \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s) \cdot I(\tau_t = s)$$

We finish the description of the model by specifying the prior. All of the priors are chosen to be proper and informative, reflecting the knowledge of the investor before observing the data. As Geweke (2007) points out, for Gaussian mixture models, proper priors for the variance parameter is essential.

$$\boldsymbol{\mu}_s \sim N(\boldsymbol{\mu}_\beta, \mathbf{V}_\beta)$$

$$\boldsymbol{\Sigma}_s^{-1} \sim \text{Wishart}(v, \boldsymbol{\Omega})$$

$$\boldsymbol{\pi} \sim \text{Dirichlet}(\alpha_1, \alpha_2, \dots, \alpha_S)$$

$\boldsymbol{\mu}_\beta, \mathbf{V}_\beta, v, \boldsymbol{\Omega}, \alpha_1, \alpha_2, \dots, \alpha_S$ are hyperparameters in the model. Though assigning different hyperparameters to $\boldsymbol{\mu}_s$ and $\boldsymbol{\Sigma}_s^{-1}$ for each latent regime is straightforward, we do not do it for notational simplicity.

3.2 Gibbs sampler

We fit the Gaussian mixtures model via Gibbs sampler, where draws from the joint posterior are taken by cycling through conditional posterior distribution. Below we outline the steps to implement the posterior simulator.

$$p(\boldsymbol{\theta}, \boldsymbol{\tau} | \mathbf{R}^h) \propto p(\boldsymbol{\theta}) \cdot \prod_{t=1}^T p(\tau_t | \boldsymbol{\theta}) \cdot p(\mathbf{R}_t | \boldsymbol{\theta}, \tau_t)$$

$$\text{Denote } \boldsymbol{\mu} \equiv \{\boldsymbol{\mu}_s\}_{s=1}^S, \boldsymbol{\Sigma} \equiv \{\boldsymbol{\Sigma}_s\}_{s=1}^S$$

Step 1 : $\boldsymbol{\mu} | \boldsymbol{\Sigma}, \boldsymbol{\tau}, \boldsymbol{\pi}, \mathbf{R}^h$

Following Lindley and Smith (1972),

$$\boldsymbol{\mu}_s | \boldsymbol{\Sigma}, \boldsymbol{\tau}, \mathbf{R}^h \sim \mathbf{N}(\mathbf{D} \mathbf{d}, \mathbf{D}), \quad \forall s = 1, 2, \dots, S$$

$$\text{where } \mathbf{D} = (\mathbf{T}_s \boldsymbol{\Sigma}_s^{-1} + \mathbf{V}_\beta^{-1})^{-1}$$

$$\mathbf{d} = \boldsymbol{\Sigma}_s^{-1} \cdot [\sum_{t=1}^T \mathbf{R}_t \cdot I(\tau_t = s)] + \mathbf{V}_\beta^{-1} \boldsymbol{\mu}_\beta$$

$$\mathbf{T}_s = \sum_{t=1}^T I(\tau_t = s)$$

Step 2 : $\boldsymbol{\Sigma} | \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\pi}, \mathbf{R}^h$

Following the standard results of Gibbs sampler in SUR model,

$$\boldsymbol{\Sigma}_s^{-1} | \boldsymbol{\mu}, \boldsymbol{\tau}, \mathbf{R}^h \sim \text{Wishart}\{v + \mathbf{T}_s, [\boldsymbol{\Omega}^{-1} + \sum_{t=1}^T (\mathbf{R}_t - \boldsymbol{\mu}_s) (\mathbf{R}_t - \boldsymbol{\mu}_s)' \cdot I(\tau_t = s)]^{-1}\},$$

$\forall s = 1, 2, \dots, S$

Step 3: $\boldsymbol{\tau} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}, \mathbf{R}^h$

The conditional posterior $\tau_t | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{R}^h$ ($\forall t = 1, 2, \dots, T$) is a discrete random variable with p.m.f.

$$\text{pr}(\tau_t = s | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{R}^h) = \frac{\pi_s \cdot \phi(\mathbf{R}_t; \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)}{\sum_{k=1}^S \pi_k \cdot \phi(\mathbf{R}_t; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}$$

$$\forall s = 1, 2, \dots, S$$

Step 4: $\boldsymbol{\pi} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\tau}, \mathbf{R}^h$

The Dirichlet prior of $\boldsymbol{\pi}$ will induce a conjugate conditional posterior distribution $\boldsymbol{\pi} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\tau}, \mathbf{R}^h$

$$\boldsymbol{\pi} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\tau}, \mathbf{R}^h \sim \text{Dirichlet}(\alpha_1 + T_1, \alpha_2 + T_2, \dots, \alpha_S + T_S)$$

With the full set of conditional posteriors, a posterior simulator proceeds by successively draws from those conditional posteriors in a cyclical fashion.

Before we move on, we probably need to address one practical issue when dealing with asset returns data, namely unequal histories of data. Due to split and merge, or whatever reasons, some stocks might have shorter observations than others. Fortunately, that complication does not lend too much difficulty to our simulator. We simply treat the missing data as a latent variable.

Without loss of generality, let us assume that in period t , the first N_1 stocks have no record of returns, while the rest $N_2 = N - N_1$ stocks have data as usual. Denote the latent N_1 stock returns by $\mathbf{R}_{t,1}^*$. Let $\mathbf{R}_t^* \equiv (\mathbf{R}_{t,1}^*, \mathbf{R}_{t,2})'$

To run the Gibbs sampling procedure, Step 1 -4 can be carried out as usual except for replacing all the \mathbf{R}_t with \mathbf{R}_t^* . Furthermore, an extra step 5 will be added.

Step 5: $\mathbf{R}_{t,1}^* | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\tau}, \boldsymbol{\pi}, \mathbf{R}^h$

Suppose $\tau_t = s$, Partition $\boldsymbol{\mu}_s = \begin{pmatrix} \boldsymbol{\mu}_{s,1} \\ \boldsymbol{\mu}_{s,2} \end{pmatrix}$, $\boldsymbol{\Sigma}_s = \begin{pmatrix} \boldsymbol{\Sigma}_{s,11} & \boldsymbol{\Sigma}_{s,12} \\ \boldsymbol{\Sigma}_{s,21} & \boldsymbol{\Sigma}_{s,22} \end{pmatrix}$ conformable with N_1, N_2

$$\mathbf{R}_{t,1}^* | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\tau}, \boldsymbol{\pi}, \mathbf{R}^h \sim \mathbf{N} \left[\boldsymbol{\mu}_{s,1} + \boldsymbol{\Sigma}_{s,12} \boldsymbol{\Sigma}_{s,22}^{-1} (\mathbf{R}_{t,2} - \boldsymbol{\mu}_{s,2}), \boldsymbol{\Sigma}_{s,11} - \boldsymbol{\Sigma}_{s,12} \boldsymbol{\Sigma}_{s,22}^{-1} \boldsymbol{\Sigma}_{s,21} \right]$$

3.3 Posterior predictive density

First note that

$$p(\mathbf{R}_{T+1}, \boldsymbol{\theta}, \boldsymbol{\tau} | \mathbf{R}^h) = p(\boldsymbol{\theta} | \mathbf{R}^h) \cdot p(\boldsymbol{\tau}_{T+1} | \boldsymbol{\theta}, \mathbf{R}^h) \cdot p(\mathbf{R}_{T+1} | \boldsymbol{\theta}, \boldsymbol{\tau}_{T+1}, \mathbf{R}^h)$$

After running the Gibbs sampler, we already have Q draws from the posterior $p(\boldsymbol{\theta} | \mathbf{R}^h)$:

$$\{\boldsymbol{\mu}^q, \boldsymbol{\Sigma}^q, \boldsymbol{\tau}^q, \boldsymbol{\pi}^q\}_{q=1}^Q$$

We can use Method of Composition to generate draws from the predictive density.

Step 1: for a given q , draw τ_{T+1}^q from p.m.f. $\text{pr}(\tau_{T+1}^q = s) = \pi_s^q$

Step 2: conditional on the value of τ_{T+1}^q , draw $\mathbf{R}_{T+1}^q | (\tau_{T+1}^q = s) \sim N(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)$

Step 3: repeat step 1-2 for $q = 1, 2, \dots, Q$, then we have $\{\mathbf{R}_{T+1}^q\}_{q=1}^Q$

Note that in our baseline model, the data generating process for \mathbf{R}_t is not history dependent—in each period, the \mathbf{R}_t comes from one of the S model $N(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)$. The implication is that if the investor plans two periods ahead, the predictive distribution for \mathbf{R}_{T+2} will be the same as that of \mathbf{R}_{T+1} , and therefore the numerical procedure to obtain draws $\{\mathbf{R}_{T+2}^q\}_{q=1}^Q$ will also be identical.

3.4 Optimal portfolio

At this stage, with simulated future returns in hand, the Markowitz mean-variance frontier can be readily constructed.

$$\text{Denote } \bar{\mathbf{R}} \equiv \sum_{q=1}^Q \mathbf{R}_{T+1}^q, \bar{\mathbf{V}} \equiv \sum_{q=1}^Q (\mathbf{R}_{T+1}^q - \bar{\mathbf{R}})(\mathbf{R}_{T+1}^q - \bar{\mathbf{R}})'$$

If we assume the investment goal is simply choose a portfolio weight $\boldsymbol{\omega}$ to minimize portfolio variance for any given expected return \bar{R}_p .

$$\sigma_p^2 \equiv \min_{\boldsymbol{\omega}} \boldsymbol{\omega}' \bar{\mathbf{V}} \boldsymbol{\omega}$$

$$\text{s.t. } \boldsymbol{\omega}' \bar{\mathbf{R}} = \bar{R}_p$$

$$\boldsymbol{\omega}' \mathbf{1} = 1$$

The solution to the portfolio optimization problem is:

$$\boldsymbol{\omega}^* = \frac{\xi \bar{\mathbf{V}}^{-1} \mathbf{1} - \alpha \bar{\mathbf{V}}^{-1} \bar{\mathbf{R}}}{\xi \delta - \alpha^2} + \frac{\delta \bar{\mathbf{V}}^{-1} \bar{\mathbf{R}} - \alpha \bar{\mathbf{V}}^{-1} \mathbf{1}}{\xi \delta - \alpha^2} \bar{R}_p$$

$$\text{Where } \xi \equiv \bar{\mathbf{R}}' \bar{\mathbf{V}}^{-1} \bar{\mathbf{R}}, \alpha \equiv \mathbf{1}' \bar{\mathbf{V}}^{-1} \bar{\mathbf{R}}, \delta \equiv \mathbf{1}' \bar{\mathbf{V}}^{-1} \mathbf{1}$$

It is very important to note that the method above is not the “certainty equivalent solution”. On the surface, we plug the sample analog of the posterior distribution into the classic mean-variance formula. However, the material difference is that the sample size Q is determined by the researcher rather than the number of observations in the data. With abundant computational resources, the sample analog ideally could infinitely get close to the population moment by increasing the repetitions of Monte Carlo experiment.

Of course, as is mentioned in the introduction, the mean-variance frontier may not be directly relevant to the investor who maximize expected utility, when the returns are not normally distributed. Therefore, it makes more sense to directly work on the maximization of expected utility.

$$\max_{\omega} \frac{1}{Q} \sum_{q=1}^Q u(\omega' \mathbf{R}_{T+1}^q)$$

$$\text{s.t. } \omega' \mathbf{1} = 1$$

Again, using the sample analog in place of the expectation operator does not imply certainty equivalent solution.

We illustrate our approach by assuming that investors have a CARA utility.

$$u(x) = -\frac{1}{\gamma} e^{-\gamma x}$$

4. Data description

To illustrate our approach, we consider a portfolio manager who diversifies investments globally. Assume she allocates funds in 9 major world stock indexes: SP500 (USA), FTSE (Britain), CAC (France), DAX (Germany), ATX (Austria), SSE (CHN), HSI (Hong Kong), NIKKEI 225 (Japan), STI (Singapore).

Daily data ranging Jan. 2000 - Jan. 2009 are used to estimate the asset returns performance.

Table 1 descriptive statistics of daily returns (%)

	USA	GBR	FRA	GER	AUT	CHN	HK	JPN	SGP
Mean	-0.024	-0.022	-0.029	-0.019	0.015	0.014	-0.010	-0.036	-0.017
Skewness	-0.130	-0.076	0.035	0.052	-0.383	-0.142	-0.327	-0.420	-1.111
Kurtosis	11.769	9.802	8.614	8.038	14.960	8.775	14.254	10.777	15.904
Lilliefors	0.083	0.081	0.072	0.071	0.107	0.087	0.095	0.072	0.079
p	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Jarque-Bera	7282.6	4380.6	2982.2	2402.9	13589.9	3163.3	12024.0	5790.3	16223.4
p	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

Table 2 sample covariance matrix of daily asset returns

	USA	GBR	FRA	GER	AUT	CHN	HK	JPN	SGP
USA	1.862	0.911	1.135	1.363	0.612	0.018	0.452	0.264	0.379
GBR	0.911	1.772	1.857	1.748	1.101	0.144	0.858	0.712	0.695
FRA	1.135	1.857	2.504	2.309	1.267	0.170	1.009	0.837	0.835
GER	1.363	1.748	2.309	2.826	1.192	0.195	0.985	0.772	0.789
AUT	0.612	1.101	1.267	1.192	1.941	0.313	0.940	0.858	0.744
CHN	0.018	0.144	0.170	0.195	0.313	2.882	0.869	0.496	0.402
HK	0.452	0.858	1.009	0.985	0.940	0.869	2.809	1.579	1.497
JPN	0.264	0.712	0.837	0.772	0.858	0.496	1.579	2.584	1.095
SGP	0.379	0.695	0.835	0.789	0.744	0.402	1.497	1.095	1.796

Table 1 and 2 provides the descriptive statistics of our dataset. Sample moments are calculated for entire length of the data from 2000 to 2009. The mean of index returns for most countries are negative, largely due to the recent financial crisis and global stock market sluggishness since 2008. The covariance matrix of returns suggests stronger positive correlations among western countries. The correlations between western and oriental markets are less prominent, which carries implications for global investment diversification.

As is seen in Table 1, in all countries, daily returns of stock indexes exhibit substantial skewness and distinct leptokurtosis. Index returns of France and Germany shows positive skewness, while others have negative skewness. The kurtosis of returns ranges from 8 to 16, with the tail heavier than the normal distribution. The Lilliefors tests (Kolmogorov-Smirnov test with estimated parameters) and Jarque-Bera test provides strong evidence against normality

with p value smaller than 0.001. Apparently, index returns of all countries deviate from normality to a large extent.

The departure from normality can also be seen from the Bayesian residual test. We first fit the model under normality assumption via Gibbs sampler. The results are shown in Table 3. The sampling procedure is essentially a simplified version described in the previous section, except that we allow only one regime. So $p(\mathbf{R}_t|\boldsymbol{\theta})$ is normally distributed here. Of course, under this simple circumstance, the analytical posterior distribution and posterior predictive distribution do exist, and should be multivariate Student-t distributed with the usual prior setup. We fit the model by Gibbs sampler with the purpose of conducting a (series of) residual test. Since we have obtained draws from the posterior sampler, i.e. $\{\boldsymbol{\mu}^q, \boldsymbol{\Sigma}^q\}_{q=1}^Q$, we use each $\boldsymbol{\mu}^q, \boldsymbol{\Sigma}^q$ to normalize the historical returns. If the returns are indeed normally distributed, then the classical Kolmogorov-Smirnov test should accept the null. The test is repeated for Q times, and the test statistics are recorded in the histogram depicted in Figure 1. The 9 panels correspond to 9 assets in sequence. Since we have a fairly large sample size of more than 2000 observations, the 1% significance critical value of Kolmogorov test can be approximated by $1.63 / \sqrt{T}$. As we can see from Figure 1, test statistics roughly center around 0.08 and almost in every circumstance the test statistics are larger than 0.06, which apparently is too large to accommodate normality.

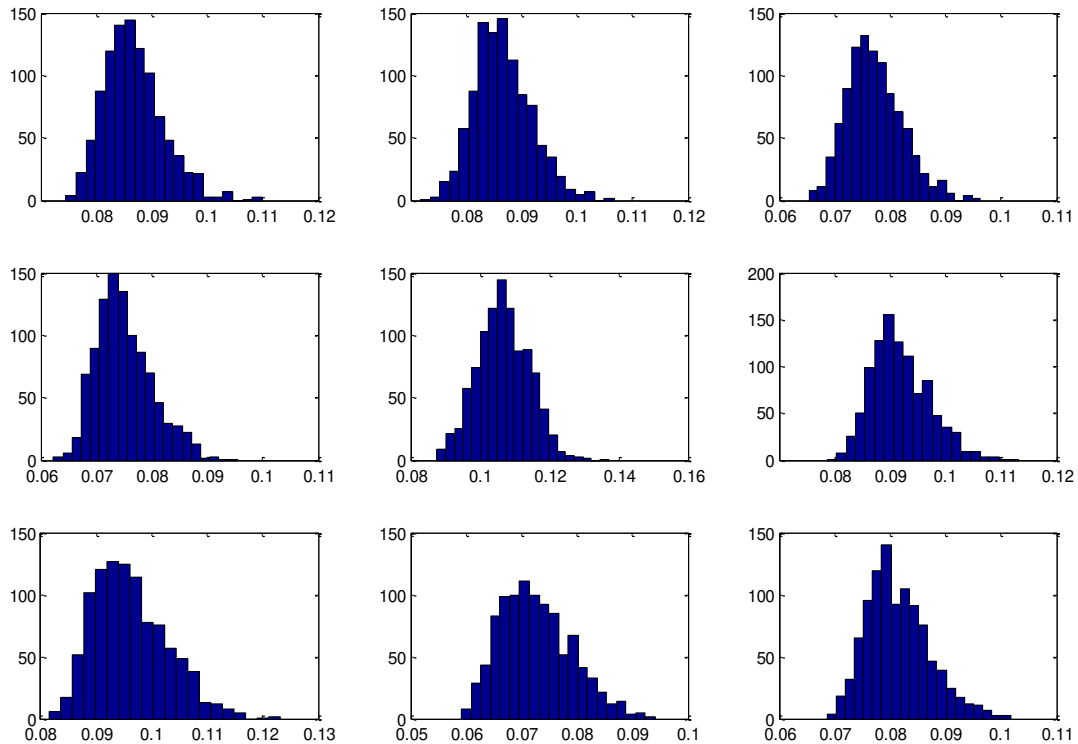
We also go through the above tests for subsamples of the data set. Without an exception, normality of asset returns are rejected. For brevity, we did not report the results in the text.

It should be noted that all the normality test are conducted on the basis of individual index returns. Once the normality is rejected at individual level, the joint normality is automatically rejected. (But not vice versa). We therefore conclude that there is a necessity to adopt some more flexible distributions to model asset returns. Finite Gaussian mixture is a good candidate which can accommodate potential multi-modality, skewness and heavy-tailness.

Table 3 Bayesian estimation under normality assumption

	USA	GBR	FRA	GER	AUT	CHN	HK	JPN	SGP
Post. Mean	-0.024	-0.022	-0.029	-0.019	0.015	0.013	-0.010	-0.036	-0.017
Post. Std	0.028	0.027	0.032	0.034	0.030	0.035	0.035	0.033	0.027
Pred. Mean	-0.023	-0.018	-0.089	-0.077	0.031	0.010	-0.027	-0.061	-0.012
Pred. Std	1.358	1.310	1.527	1.643	1.344	1.730	1.698	1.609	1.361

Figure 1 Bayesian residual Kolmogorov-Smirnov test statistics



5. Bayesian portfolio selection

In this section, we discuss the procedure of a typical Bayesian portfolio selection. Suppose the portfolio manager uses historical returns from 2000-2009 as the input of her decision making. On the basis of her forecast of future stock returns, she optimally allocates the portfolio weights in each market.

As is pointed out in the previous section, the world-wide stock markets crash in 2008 rendered negative returns of most stocks, many of which lose half of its market value. In hindsight, it would have been better to lock the money in the coffer, rather than invest any dollar in the stock market. For illustration purposes, we exclude the possibility of refraining from investment and assume no safe assets. We intentionally include the returns in the sagging period as well as the booming and leveling-off periods, because it is a natural experiment of several regimes.

We use a three-regime Gaussian mixtures model to describe the stock returns.

Table 4 reports the posterior distribution of index returns in each regime, and Table 5 shows the posterior predictive distribution. In this round of simulation, the three regimes occur with probability 0.09, 0.46, 0.45 respectively. It seems that Regime 1 resembles the regime of “crash”; Regime 2 represent some leveling off; and Regime 3 is the booming case. However, due to the complication induced by the permutation problems, the above result should be interpreted with a grain of salt.

Nevertheless, the posterior predictive distribution is invariant with respect to permutation, so that it can be interpreted without ambiguity. Comparing Table 4 with the descriptive statistics in Table 1, we find that Bayesian forecast of future returns are reasonably close to, but not the same as, the sample moments of historical returns. The difference is due to the explicit account for the uncertainty as well as nonnormality of our Bayesian approach.

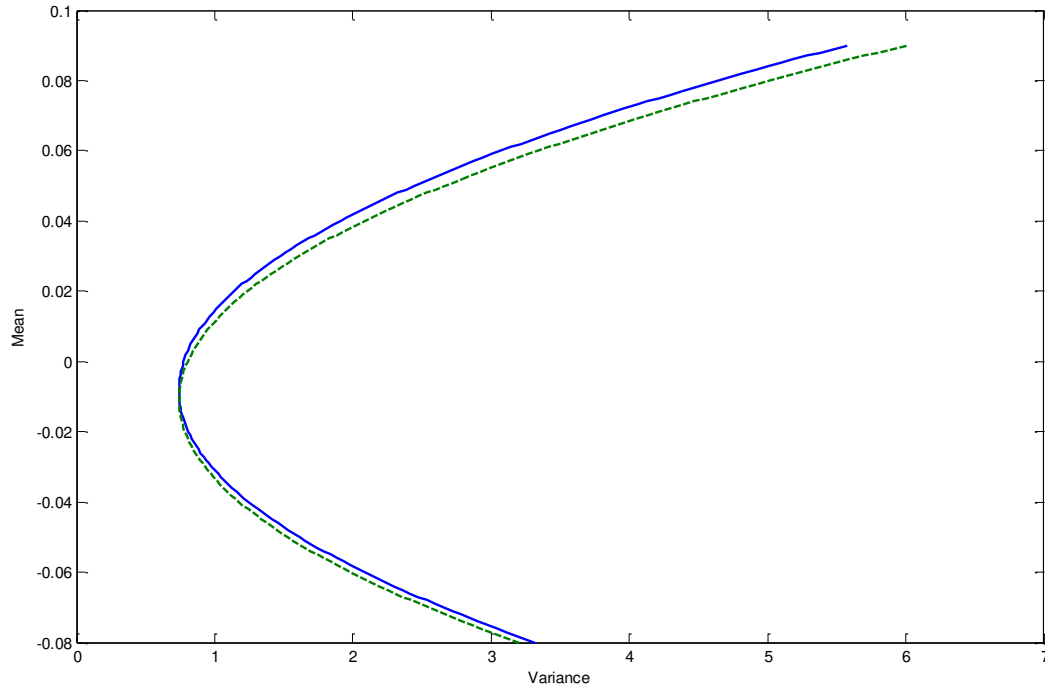
Table 4 Posterior distribution of each regime

	USA	GBR	FRA	GER	AUT	CHN	HK	JPN	SGP
Regime 1									
Post.mean	-0.131	-0.258	-0.250	-0.245	-0.548	-0.379	-0.392	-0.397	-0.412
Post.std	0.223	0.212	0.237	0.233	0.248	0.261	0.283	0.248	0.217
Regime 2									
Post.mean	-0.068	-0.075	-0.105	-0.104	-0.067	-0.043	-0.076	-0.120	-0.061
Post.std	0.046	0.047	0.057	0.064	0.040	0.044	0.054	0.053	0.045
Regime 3									
Post.mean	0.044	0.081	0.097	0.118	0.214	0.153	0.137	0.127	0.110
Post.std	0.024	0.022	0.027	0.030	0.029	0.054	0.031	0.037	0.026

Table 5 Posterior predictive distribution of index returns (%)

	Mean	Std	Pr(R>0)
USA	-0.022	1.360	0.501
GBR	-0.020	1.323	0.511
FRA	-0.026	1.573	0.511
GER	-0.014	1.674	0.514
AUT	0.020	1.384	0.532
CHN	0.013	1.693	0.507
HK	-0.003	1.672	0.517
JPN	-0.029	1.599	0.506
SGP	-0.011	1.333	0.516

Figure 2 Bayesian and CE mean-variance frontier



Real line represents Bayesian Mean-Variance frontier

Dotted line represents certainty equivalent MV frontier

Using draws from posterior predictive distribution, Figure 2 depicts the Bayesian mean-variance frontier. For comparison, we also provide the mean-variance frontier with certainty equivalent approach. Again the two curves are close to each other, but Bayesian frontier predicts a slightly higher expected return with a given variance.

Mean-variance frontier may not be directly relevant to decision making in the presence of nonnormality. The next step is to estimate the optimal portfolio weights which maximize the expected utility. We assume that the portfolio manager has the CARA utility with absolute risk aversion coefficient γ . Table 6 shows the optimal weights with varied risk aversion coefficients. The weights sound reasonable, at least not extreme. Obviously, there is a need for short-sell to attain fully diversification. Table 7 further provides optimal weights when short sell are not allowed. In that case, the previous negative weights are replaced by zero, and portfolio shares are allocated among 4-5 indexes.

Table 6 Bayesian U-max portfolio weights (%) with varied risk aversion coefficients

	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$	$\gamma = 5$	$\gamma = 6$	$\gamma = 7$	$\gamma = 8$
USA	30.60	27.76	25.16	22.98	21.42	20.36	20.36	20.36
GBR	42.95	46.91	46.43	46.74	46.76	46.51	46.51	46.51
FRA	-22.56	-19.49	-17.49	-17.70	-18.11	-18.40	-18.40	-18.40
GER	-2.13	0.96	2.48	4.83	7.06	8.97	8.97	8.97
AUT	-0.73	-10.41	-12.40	-13.24	-13.37	-13.22	-13.22	-13.22
CHN	26.33	30.89	33.80	35.85	37.26	38.10	38.10	38.10
HK	-22.51	-30.46	-31.18	-30.52	-29.45	-28.37	-28.37	-28.37
JPN	15.49	18.88	17.36	14.71	11.93	9.61	9.61	9.61
SGP	32.57	34.95	35.84	36.35	36.50	36.43	36.43	36.43

Table 7 Bayesian U-max portfolio weights (%) , no short-sell allowed

	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$	$\gamma = 5$	$\gamma = 6$	$\gamma = 7$	$\gamma = 8$
USA	28.56	26.53	25.04	24.24	23.84	23.60	23.60	23.60
GBR	19.59	25.01	27.71	28.20	27.75	26.75	26.75	26.75
FRA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GER	0.00	0.00	0.00	0.00	0.00	0.30	0.30	0.30
AUT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CHN	22.73	23.18	24.52	25.45	26.25	27.00	27.00	27.00
HK	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
JPN	6.91	2.69	0.45	0.00	0.00	0.00	0.00	0.00
SGP	22.19	22.60	22.28	22.10	22.16	22.34	22.34	22.34

6. Simulated investment championship

In this section, we illustrate our approach by conducting an investment competition. We will compare the performance of Bayesian portfolio selection with Gaussian-mixtures, Bayesian portfolio selection with normal distribution assumption, and certainty equivalent approach.

The rules are as follows:

- i) Use a rolling window consisting of 1000 historical data for estimation, whatever methods.
- ii) Rebalance the portfolio weights every 180 business days.
- ii) Charge 0.5% transaction cost either purchase or sell stocks, including the first and last period.

The championship starts on the 1000th observation (i.e., at the end of the year 2003). Due to the simplicity of our model, returns are *ex ante* i.i.d. That is to say, forecasting tomorrow's returns is the same as that of multiple-periods ahead. With initial funding of \$10000, investment can be diversified in 9 countries, with short sell allowed. Exchange rates are assumed to be constant, so stock indexes, denominated in whatever currency, can be conceptually treated as dollar price.

Since our dataset contains 2271 observations, there are 7 rebalance opportunities. In our algorithm, transaction costs are viewed as *ex post*, i.e., when the agent optimally chooses portfolios, she does not take transaction costs into account. That assumption, of course, is only made for the sake of convenience, since transaction costs would be a constraint on the dynamic optimal portfolio in that it requires the portfolio weight should not be too volatile across periods. We introduce the *ex post* transaction cost for the purpose of evaluating the performance the three portfolio selection methods, which may differ in turnover rate.

To alleviate the computational burden, Gaussian mixtures takes only two regimes, which could naturally be interpreted as bull market and bear market. Our simple model does not introduce the regime correlations over time, though that feature could be added simply by using Markov transition of regimes.

The utility function is assumed to be CARA with risk aversion coefficient 2. For Bayesian Gaussian-mixture model, simulations are used to estimate the optimal portfolio. For Bayesian normal model and certainty equivalent model, an easier algorithm is employed. We first calculate the mean-variance frontier with fine grids of expected portfolio return. It is well-known that the CARA utility, coupled with normal returns, implies optimal portfolio can be analytically derived as minimization of:

$$E(R_p) - \frac{1}{2} \gamma \text{Var}(R_p)$$

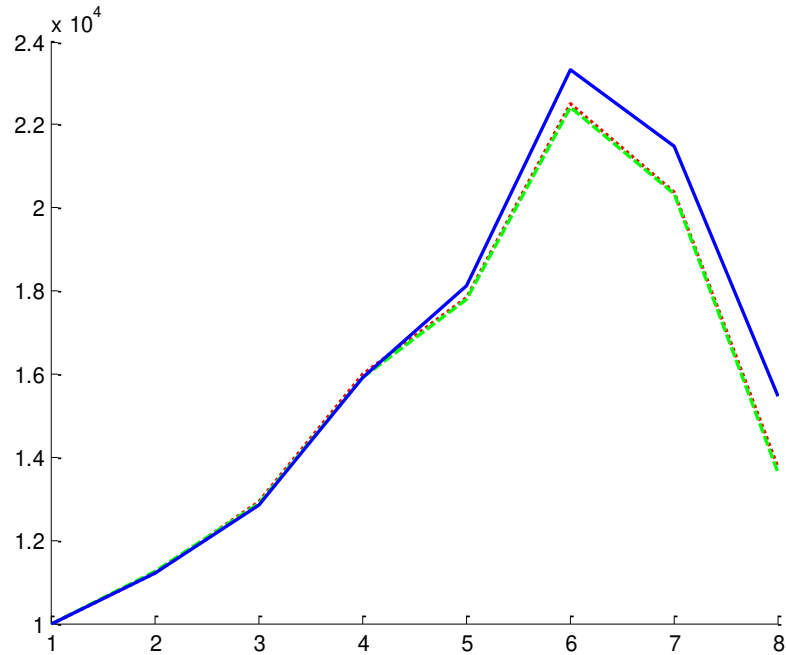
In that spirit, grid search the $E(R_p)$, $\text{Var}(R_p)$ pairs derived from the mean-variance frontier, we will obtain the optimal portfolio weights. The only less desirable fact is that in the Bayesian portfolio selection with normal returns (likelihood function), the posterior predictive distribution is actually multivariate Student-t rather than normal distribution. Since they are more or less similar in shape, we stick to the above simple and fast algorithm to obtain the optimal weights.

The simulation results are reported in Figure 3, 4 and 5. It is clearly shown that Bayesian portfolio selection with Gaussian-mixtures outperforms the other two in term of final wealth. In the first half of the investment periods (2003-2007), global market is booming in general. The wealth level under alternative methods go up neck and neck. However, the year 2008

witnessed downturn of the market. The advantage of Bayesian portfolio selection with Gaussian-mixtures is fully revealed. That is because mixture model has taken the possibility of bear market into account when the bull market is on. The portfolio should be relatively robust when the bad day suddenly falls.

Figure 5 shows the turnover rate, which is defined as the ratio of transaction amount over the total wealth, of the three strategies. By definition, in the first period the agent uses up the initial funding to buy shares. Since short sell is allowed, the turnover rate can be larger than 100%. In the last period, all the shares holding will be converted to cash in order to compare the final wealth; the turnover rate will also exceed 100% due to short selling. It can be seen from Figure 5 that Bayesian portfolio selection with Gaussian-mixtures has a higher turnover rate than the rest two. We are not very sure whether is it the feature of mixture models itself or simply induced by the additional uncertainty induced by Monte Carlo simulation. (After all, the posterior simulator of mixture models usually mix slowly, large numeric error is associated accordingly.) Be that as it may, Figure 4 indicates that Bayesian portfolio selection with Gaussian-mixtures outperforms the other two even if transaction costs are included.

Figure 3 simulated portfolio wealth over time (without transaction cost)



Blue Real line represents Bayesian portfolio selection with 2 regions

Green dashed line represents Bayesian portfolio selection with normality assumption

Red dotted line represents certainty-equivalent portfolio selection

Figure 4 simulated portfolio wealth over time (with transaction cost)

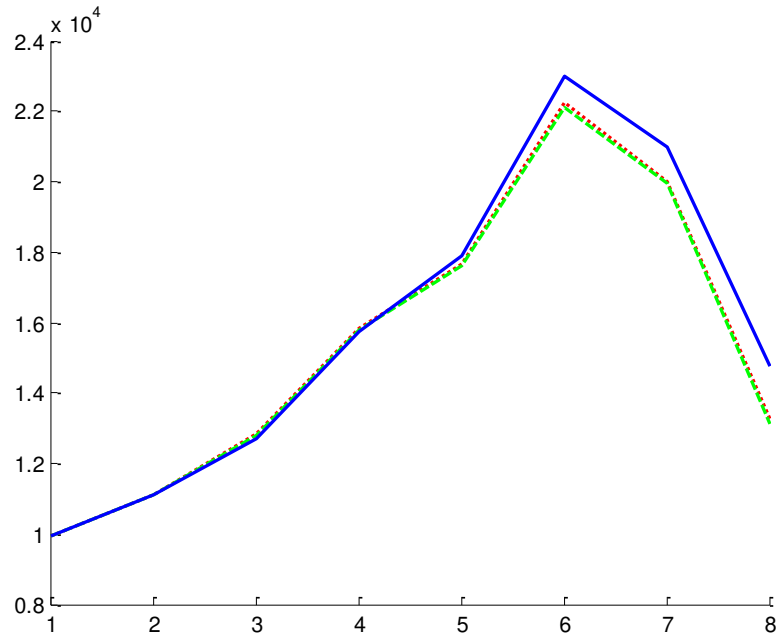
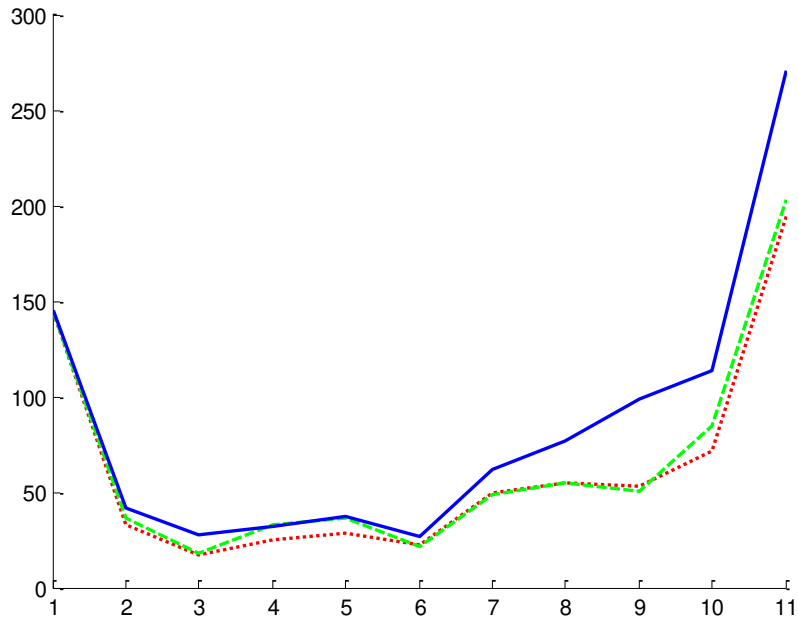


Figure 5 turnover rate under three trading strategies



Blue Real line represents Bayesian portfolio selection with 2 regions

Green dashed line represents Bayesian portfolio selection with normality assumption

Red dotted line represents certainty-equivalent portfolio selection

7. Extensions

The Gaussian-mixture Bayesian model employed above can be viewed as the basic version of a broader class of mixture models. Our baseline model could readily be extended in various lines.

Firstly, we might add more factors to increase explanatory and predictive power.

In the baseline model, in regime $s \in \{1, 2, \dots, S\}$, we have

$$\mathbf{R}_t = \boldsymbol{\mu}_s + \boldsymbol{\varepsilon}_{t,s}, \boldsymbol{\varepsilon}_{t,s} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_s)$$

There is no need to treat $\boldsymbol{\mu}_s$ as a constant. Stock returns, in some senses, are predictable. Historical returns and/or fundamentals (PE ratio, PB ratio, market return, etc) may have explanatory power over the future stock returns.

To accommodate predictability of returns, we can extend the baseline model to be:

$$\mathbf{R}_t = \mathbf{X}_t \boldsymbol{\beta}_s + \boldsymbol{\varepsilon}_{t,s}, \boldsymbol{\varepsilon}_{t,s} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_s)$$

Where \mathbf{X}_t may contain lagged value of asset returns, and exogenous indicators pertaining to the fundamentals of the assets. In that case, we can still employ the informative priors such as $\boldsymbol{\beta}_s \sim \mathbf{N}(\boldsymbol{\mu}_\beta, \mathbf{V}_\beta)$, $\boldsymbol{\Sigma}_s^{-1} \sim \text{Wishart}(v, \boldsymbol{\Omega})$. Gibbs samplers for the SUR model can be conducted in the usual way.

The explanatory factors and predictive power can be easily tested in the Bayesian framework by calculating the marginal likelihood. We might even retain the modeling uncertainty by Bayesian averaging the potential models. All the inputs we need are the (numerical) marginal likelihood of each model, which can be simulated using the Chib (1995) Method with the aid of Gibbs sampler.

Secondly, we might use more parsimonious representations of the mixture models.

Finite Gaussian mixture models is a highly flexible model which is capable of mimicking various shapes of p.d.f. Multimodality, skewness and leptokurtosis or platykurtosis are addressed in an all-in-one fashion. However, the flexibility comes at a price—the computational burden associated with the finite Gaussian mixture model is huge. For this model, Gibbs sampler is notorious for its local movement. The chain mixes very slowly and large numbers of draws need to be taken for satisfactory performance of the sampler. The reason, as Jasra et al.(2005) put it, is the inability of Gibbs sampler to traverse the multimodal density to and fro.

The descriptive statistics usually suggest the high frequency financial data exhibit heavy-tail and skewness. The former can be more compactly addressed by introducing a latent scale adjustment in the variance of the error terms. If the latent scale variable is inverted gamma distributed, it usually can lead to a Student-t distributed likelihood function so as to accommodate the fat-tail. The latter can be more parsimoniously modeled by introducing a latent location adjustment variable, which is usually assumed to be half-normally distributed. The positive adjustment term, therefore, induce the skewness to the right, while the negative one will bring in left-ward skewness.

8. Conclusion

Parameter estimation risk and non-normal asset returns are two major barriers to the implementation of Markowitz portfolio selection. This paper attempts to address the two issues in a unified Bayesian framework, in which parameter uncertainty are reflected in the posterior predictive distribution of asset returns, and deviation from normality is captured by finite Gaussian mixtures. We develop a Gibbs sampling procedure to obtain draws from the posterior distribution as well as draws from the predictive density. Then the portfolio weights can be optimally constructed so as to maximize the expected utility of investors.

To illustrate our approach, we considered a simplified version of global diversification—funds to be invested in several leading stock market indexes. The descriptive statistics provide strong evidence against normality of high frequency index returns, hence the need of mixture models. Bayesian portfolio selection with 3-region mixtures is used to predict the future returns, which tracks the data closely. The associated optimal portfolios are also reasonably diversified among assets.

A simulated investment championship is then conducted to evaluate the relative performance of three portfolio selection strategies, namely Bayesian mixtures, Bayesian normal, and certainty equivalence. It turns out that Bayesian mixtures model does outperform the rest two, especially in the periods of world-wide downturn since 2008. The prominent predictive power of Bayesian mixture is not accidental; it is rooted in its ability to taken the possibility of bear market into account when the bull prevails.

References

- Black, F., and Litterman, R., 1992. Global portfolio optimization. *Financial Analysts Journal*, 48:28–43.
- Bawa, V., Brown, S. and Klein, R. Estimation Risk and Optimal Portfolio Choice. Amsterdam: North Holland Publishing, 1979.
- Celeux, G., Hurn, M., Robert, C.P., 2000. Computational and inferential difficulties with mixture posterior distributions. *Journal of American Statistics Association*, 95, 957–970.
- Chib, S., 1995. Marginal likelihood form the Gibbs sampler. *Journal of the American Statistical Association*, 90, 1313-1321.
- Chopra, V. K., and Ziemba, W. T., 1993. The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice. *Journal of Portfolio Management*, Vol. 19, 2, 6-11.
- Dickinson, J., 1974. The reliability of estimation procedures in portfolio analysis. *Journal of Financial and Quantitative Analysis* 9, 447–462.
- Fama, E. 1965. The behavior of stock market prices. *Journal of Business*. 38. pp. 34-105.
- Frost, P. A., and Savarino, J. E. ,1986. An Empirical Bayes Approach to Efficient Portfolio Selection. *Journal of Financial and Quantitative Analysis*, 21, 293-305.
- Frühwirth-Schnatter, S., 2001. Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of American Statistics Association* ,96, 194–209.
- Jasra, A., Holmes, C.C., Stephens, D.A., 2005. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistists Science*, 20, 50–67.
- Jobson, J., Korkie, B., 1980. Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association* 75, 544–554.
- Jorion, P., 1986. Bayes–Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis* 21, 279–291.
- Geweke, J., 2007. Interpretation and inference in mixture models: simple MCMC works. *Computational Statistics and Data Analysis*, 51, 3529-3550.
- Goldfarb, D., and Iyengar, G. 2003. Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38.
- Greyserman, A., Jones, D., Strawderman, W., 2006. Portfolio selection using hierarchical Bayesian analysis and MCMC methods. *Journal of Banking & Finance* 30, 669–678.
- Harvey, C. R., and Siddique, A., 2000. Conditional Skewness in Asset Pricing Tests, *Journal of Finance*, 55, 1263-1295.
- Klein, R. W., and Bawa, V. S. ,1976. The Effect of Estimation Risk on Optimal Portfolio Choice. *Journal of Financial Economics*, 3, 215-231.
- Kon, S.J., 1984. Models of Stock Returns--A Comparison. *Journal of Finance*, Vol. 39, 1, pp. 147-165
- Kraus, A., and Litzenberger, R., 1976. Skewness Preference and the Valuation of Risk Assets, *Journal of Finance*, 31, 1085-1100.
- Lindley, D. V., Smith, A., 1972. Bayes Estimates for the Linear Model, *Journal of the Royal Statistical Society*, Vol. 34, No. 1, 1-41.
- Markowitz, H. M. , 1952. *Portfolio Selection: Efficient Diversification of Investments*, New York: Wiley.

- Newcomb, S., 1886. A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics*, 8, 343–366.
- Michaud, R. O. 1998. *Efficient Asset Management: A Practical Guide to Stock Portfolio Management and Asset Allocation*, Financial Management Association, Survey and Synthesis Series. HBS Press, Boston, MA.
- Peiro, A., 1994. The distribution of stock returns: International evidence, *Applied Financial Economics*, v. 4. pp. 431-439.
- Polson, N. G., Tew, B. V., 2000. Bayesian Portfolio Selection: An Empirical Analysis of the S&P 500 Index 1970-1996, *Journal of Business & Economic Statistics*, Vol. 18, 2, 164-173.
- Praetz, P., 1972. The distribution of share price changes, *Journal of Business*, 45, pp. 49-55.
- Rachev, S., Hsu, J., Bagasheva, B., Fabozzi, F., 2008. *Bayesian Methods in Finance*. John Wiley & Sons.
- Scherer, B. 2002. Portfolio resampling: Review and critique. *Financial Analysts Journal*, 58(6):98–109.
- Weigand, A.S., Shi, S.M., 2000. Predicting daily probability distributions of S&P 500 returns. *Journal of Forecasting* 19, 375–392.