Bayesian Portfolio Selection in a Markov Switching Gaussian Mixture Model

Qian, Hang

Iowa State University

24 December 2011

Bayesian Portfolio Selection in a Markov Switching Gaussian Mixture Model

Hang Qian[1]

Abstract

Departure from normality poses implementation barriers to the Markowitz mean-variance portfolio selection. When assets are affected by common and idiosyncratic shocks, the distribution of asset returns may exhibit Markov switching regimes and have a Gaussian mixture distribution conditional on each regime. The model is estimated in a Bayesian framework using the Gibbs sampler. An application to the global portfolio diversification is also discussed.

Keywords: Portfolio, Bayesian, Hidden Markov Model, Gaussian Mixture.

1. Introduction

Markowitz (1952) mean-variance analysis is consistent with an investor's utility maximization when the asset returns are normally distributed. However, it has long been recognized by practitioners and scholars that financial asset returns often depart from normality. Investors usually feel that the stock prices crawl upwards for months but plummet in a day, which is hard to reconcile with the relative thin and symmetric tail of a normal distribu-

[1]Corresponding author: Department of Economics, Iowa State University, Ames, IA, 50010. Email: hqi@iastate.edu

tion. There is abundant of empirical evidence suggesting unconditional asset returns exhibit skewness, fat fail and extreme values (e.g., Fama, 1965; Blattberg and Gonedes, 1974; Peiro, 1999; Ane and Geman, 2000, to name a few). In that case, the mean and variance are inadequate to characterize all the relevant aspects of the optimal portfolio. Higher order moments will play a role in the portfolio selection and asset pricing. From a theoretical perspective, Kraus and Litzenberger (1976) model the skewness preference and its effect on risk assets valuation. Harvey and Siddique (2000) empirically test the effect of conditional skewness on the asset pricing. Jondeau and Rockinger (2006) use a fourth order Taylor expansion of the expected utility to quantify the extent to which non-normality affects the optimal asset allocation.

Given the stylized fact of departure from normality, it is natural to propose some other distributions that can better accommodate skewness, leptokurtosis and extreme values. Some early works use symmetric stable distribution (Fama, 1965) and Student-t distribution (Blattberg and Gonedes, 1974) to account for the fat tail but not skewness. Harvey et al. (2010) consider the portfolio selection with a multivariate skew normal distribution, which allows skewness and coskewness. A skew normal random variable is essentially the sum of a normal and half normal variate. It is not easy to provide an economic interpretation of the half normal variate, since the sum of two half normal is no longer half normal. Modeling asset returns with skew normal distribution is largely an empirical strategy that conveniently and effectively addresses the concern of skewed returns. Buckley et al. (2008) use the Gaussian mixture distribution in the portfolio optimization. Gaussian mixture has the natural interpretation of, say, distressed and tranquil market regimes. Furthermore, Gaussian mixture can mimic many complicated distribution. For instance, if we assign a large probability to the first regime and a low probability to the second regime with a large mean and large variance, the mixture is similar to a positively skewed distribution. If the second regime has a large mean and small variance, the mixture is close to a normal distribution with occasional extreme values. If we mix infinite number of normal distributions with the same mean and inverse-gamma distributed variances, the mixture is a Student t distribution that allows leptokurtosis.

In the Gaussian mixture, the latent regimes are random draws from a multinomial distribution without autocorrelation. As a generalization, if the latent regimes are allowed to have a Markov law of motion, the mixture becomes a hidden Markov model (HMM) initially proposed by Baum and his colleagues (Baum and Petrie, 1966; Baum and Eagon, 1967; Baum et al., 1970). It has been successfully applied to a variety of fields such as speech recognition, signal process (Rabiner, 1989). This model is mostly known to economists in another name: the Markov (regime) switching model. Hamilton (1989) models the mean GNP growth rate with two Markov switching regimes. Turner et al. (1989) consider the variance of a portfolio's excess returns depends on a Markov switching state variable. There are many subsequent works that extend the Markov switching model to vector autoregression (Krolzig, 1997), endogenous regime transition (Kim et al., 2008), etc.

In this paper, we consider a scenario that asset prices are affected by two types of latent shocks. The first type is common shocks on all assets, which lead to Markov switching regimes. The second type is idiosyncratic shocks on each individual asset, which lead to a Gaussian mixture returns with different states. The joint forces of the common and idiosyncratic shocks make asset returns follow a Markov switching Gaussian mixture (MSGM) distribution.

It is possible to write down the likelihood function of our model in a recursive manner, and then estimate the model by maximum likelihood or E-M algorithm. However, when the number of regimes/states become moderately large, say more than three, the model contains many parameters and the numerical maximization can hardly perform satisfactorily in practice. [2] Roman et al. (2010) conclude that "although theoretically the HMM-based time series modeling can tackle the multivariate case, from a practical point of view it has limited applicability". Our model is estimated in a Bayesian framework. Given an appropriate estimation routine, it can reliably estimate model for moderately large number of regimes/states with affordable

[2]There are many available estimation routines of Markov switching model. On his website, Professor James Hamilton collects links to these programs. A well-received MATLAB program by Marcelo Perlin estimates the model by direct maximization of the log likelihood. In the manual, Marcelo advises against the use of the model with more than three regimes. He mentions "the solution is probably a local maximum and you can't really trust the output you get". In the MATLAB statistic toolbox there is a routine "hmmtrain.m"for estimating discrete HMM model by the E-M algorithm. However, we test the routine with three or four regimes using pseudo data with known data generating process, the program cannot reliably estimate the model if the initial values are not carefully chosen. In a tutorial to regime-switching models, Hamilton (2005) mentions most applications assume two or three regimes and "there is considerable promise in models with a much larger number of regimes, either by tightly parameterizing the relation between the regimes (Calvet and Fisher, 2004), or with prior Bayesian information (Sims and Zha, 2006) ".

computation costs.

Estimating our model in the Bayesian framework has another advantage. The classic Markowitz portfolio selection has an implementation barrier called the "estimation risk", that is, our inability to provide the exact inputs since the population mean and covariance matrix of asset returns are unknown. One might resort to the sample analog as a certainty equivalent solution. It is well documented that the portfolio weights under that approach tend to be volatile, sensitive to minor inputs change and lack diversification (Dickinson, 1974; Jobson and Korkie, 1980; Black and Litterman, 1992; Michaud and Michaud, 2008, among others). The instability of portfolio weights is believed to be caused by the negligence of parameter uncertainty, especially the estimation error of the mean. Chopra and Ziemba (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 will be assigned larger weight in the portfolio, but those asset returns are more likely to be error-ridden (Scherer, 2002). In the Bayesian framework, the problem of the estimation risk is resolved since our triple uncertainties over the regimes/states, parameters and future disturbances are fully embodied in the posterior predictive distribution.

The rest of the paper is organized as follows. Section 2 lays out a microfoundation that the assets prices and returns follow the MSGM distribution when the assets are subject to common and idiosyncratic shocks. Section 3 outlines the econometric model and the Gibbs sampler to obtain posterior draws of model parameters as well as hidden regimes and states. Section 4 discusses the optimal portfolio selection using the posterior draws of future asset returns. Section 5 provides an illustrative application and compares performance our model to the classic portfolio selection model. Section 6 concludes the paper.

2. Micro Foundation

Though our asset returns model is primarily an empirical econometric model that flexibly accommodates non-normality, it has a micro foundation. The hidden Markov Gaussian mixture asset returns can be justified by a Lucas asset-pricing model (Lucas, 1978), where we slightly adapt the classic model by decomposing the exogenous productivity shocks into common and idiosyncratic components. The former induces Markov switching regimes and the latter leads to mixture normal returns conditional on a regime.

Consider a pure exchange economy consisting of numerous identical agents with n varieties of fruit trees that are symbolic of assets. Normalize the number of agents and trees of each variety to one. At the beginning of a period, trees yield stochastic fruits, a parable of exogenous productivity shocks. Then agents eat fruits and trade trees at market prices. Assume logarithmic preferences, the sequential optimization problem of an agent is formulated as

$$\max_{\{C_{i,t},A_{i,t}\}} E_0 \left[\sum_{t=0}^{\infty} \sum_{i=1}^{n} \beta^t \ln C_{i,t} \right],$$

$$\sum_{i=1}^{n} (C_{i,t} + p_{i,t} A_{i,t+1}) = \sum_{i=1}^{n} (p_{i,t} A_{i,t} + d_{i,t} A_{i,t}),$$

where $d_{i,t}$ is the yield of fruit i in period t. At the market price $p_{i,t}$, an agent holds $A_{i,t+1}$ amount of trees into the next period and consumes $C_{i,t}$. Assume that exogenous productivity shocks are mainly determined by the weather dummy w_t and vermin dummy $v_{i,t}$ in a regression style:

$$\ln d_{i,t} = \alpha_i + \gamma_i \cdot w_t + \delta_i \cdot v_{i,t} + \varepsilon_{i,t}.$$

The baseline yield of fruit i is $\exp(\alpha_i)$. The period-t weather condition w_t simultaneously affects trees of all varieties, though with varied magnitudes on different fruits. Assume w_t follows a Markov chain with two regimes. The weather condition is symbolic of common shocks in the macroeconomy that affect all financial assets. As the weather can be "sun" or "rain", the market can be heuristically labeled as "bull" or "bear". On the other hand, trees may be also subject to vermin intrusion and assume each variety of fruit tree is vulnerable to a species of worm, captured by the dummy variable $v_{i,t}$ for fruit i in period t. The presence of vermin is a metaphor of major idiosyncratic shocks to a financial asset. Lastly, the disturbance $\varepsilon_{i,t}$ captures countless minor common or idiosyncratic factors that may affect fruits harvest. Assume $(\varepsilon_{1,t}, ..., \varepsilon_{n,t} | w_t = s) \sim N(\mathbf{0}, \mathbf{\Sigma}_s)$, $s = 1, 2$.

The solution to the Lucas tree model is standard. By iterating forward the Euler equation

$$p_{i,t} = E_t \left[\beta \frac{C_{i,t}}{C_{i,t+1}} (p_{i,t+1} + d_{i,t+1}) \right], \ i = 1, ..., n, \ t \geq 0,$$

we obtain the fundamental asset price without bubbles

$$p_{i,t} = E_t \left[\sum_{j=1}^{\infty} \beta^j \frac{C_{i,t}}{C_{i,t+j}} d_{i,t+j} \right].$$

On the other hand, the market clearing condition requires $C_{i,t+j} = d_{i,t+j}$, $\forall j \geq 0$. So the asset pricing equation is eventually given by

$$p_{i,t} = \frac{\beta}{1-\beta} d_{i,t},$$

or

$$\ln p_{i,t} = [\ln \beta - \ln (1-\beta) + \alpha_i] + \gamma_i w_t + \delta_i v_{i,t} + \varepsilon_{i,t}.$$

The asset return from period $t-1$ to t consists of capital gains and dividend income, which can be thought as the log difference of the dividend-adjusted price series.

$$\begin{aligned} r_{i,t} &\equiv \ln (p_{i,t} + d_{i,t}) - \ln p_{i,t-1} \\ &= -\ln \beta + \gamma_i (w_t - w_{t-1}) + \delta_i (v_{i,t} - v_{i,t-1}) + (\varepsilon_{i,t} - \varepsilon_{i,t-1}) \end{aligned}$$

Typically a researcher observes neither common nor idiosyncratic shocks, therefore we marginalize $\ln p_{i,t}$ and $r_{i,t}$ with respect to w_t, $v_{i,t}$. First consider the distribution of logarithmic asset prices. $(\ln p_{1,t}, ..., \ln p_{n,t})$ follows a two-regime hidden Markov chain. Conditional on each regime, there are 2^n latent states determined by idiosyncratic shocks. Further conditional on each state, it follows a multivariate normal distribution. In other words, the distribution of logarithmic asset prices is a hidden Markov Gaussian mixture. Next consider the distribution of asset returns. Put the weather conditions in period t and period $t-1$ as a pair, the joint returns $(r_{1,t}, ..., r_{n,t})$ follows a four-regime hidden Markov chain. Conditional on each regime, it is a Gaussian mixture with 4^n latent states. Therefore, the Lucas tree model justifies both logarithmic asset prices and returns follow the MSGM distribution.

3. Econometric Model

In this section, we build an econometric model of asset prices determination in the spirit of the theoretical model outlined in Section 2.

Let the returns of n assets be $\mathbf{Y}_t = (r_{1,t}, ..., r_{n,t})', t = 1, ..., T$. Denote $\mathbf{Y}_1^T = \{\mathbf{Y}_t\}_{t=1}^T$. Assume assets returns are driven by a hidden Markov chain with S regimes. Let the latent regime in period t be $\tau_t \in \{1, ..., S\}$ and denote $\tau_1^T = \{\tau_t\}_{t=1}^T$. The initial (period 1) distribution is $\pi = (\pi_1, ..., \pi_S)'$ and the transition matrix is given by

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_1 \\ ... \\ \mathbf{Q}_S \end{pmatrix} = \begin{pmatrix} Q_{1,1} & ... & Q_{1,S} \\ ... & & \\ Q_{S,1} & ... & Q_{S,S} \end{pmatrix}.$$

Conditional on τ_t, $\mathbf{Y}_t$ follow a Gaussian mixture with K latent states. Let the latent states be $\lambda_t \in \{1, ..., K\}$, following a multinomial distribution with probability $\eta_s = (\eta_{s,1}, ..., \eta_{s,K})'$ under the current regime s. If τ_t, λ_t were known, $\mathbf{Y}_t$ would be a multivariate normal vector.

$$P(\mathbf{Y}_t | \tau_t, \lambda_t) = \prod_{s=1}^{S} \prod_{k=1}^{K} [\phi(\mathbf{Y}_t; \mu_{s,k}, \mathbf{\Sigma}_{s,k})]^{I(\tau_t = s) \cdot I(\lambda_t = k)},$$

where $\phi(\cdot)$ is the density of a multivariate normal distribution, and $I(\cdot)$ is an indicator function that takes one if the expression in the parenthesis is true and takes zero otherwise. Note that the theoretic model in Section 2 suggests the covariance matrix $\mathbf{\Sigma}_{s,k}$ does not change with the latent state k. This restriction can greatly reduce the number of parameters in the model, though as a general econometric model we allow the covariance matrix to vary with regimes and states.

Conjugate proper priors of model parameters are specified as

$$\begin{aligned}
\mu_{s,k} &\sim N\left(\mathbf{b}_{s,k}, \mathbf{V}_{s,k}\right), \\
\left(\mathbf{\Sigma}_{s,k}\right)^{-1} &\sim Whishart\left(\mathbf{\Omega}_{s,k}, \nu_{s,k}\right), \\
\mathbf{Q}_s &\sim Dirichlet\left(a_{s,1}, ..., a_{s,S}\right), \\
\pi &\sim Dirichlet\left(c_1, ..., c_S\right), \\
\eta_s &\sim Dirichlet\left(f_{s,1}, ..., f_{s,K}\right),
\end{aligned}$$

where $s = 1, ..., S$ and $k = 1, ..., K$.

The full posterior conditional distribution of $\mu_{s,k}$ is given by

$$\mu_{s,k} \,|\cdot \sim N\left(\mathbf{D}_{s,k}\mathbf{d}_{s,k}, \mathbf{D}_{s,k}\right),$$

where

$$\begin{aligned}
\mathbf{D}_{s,k} &= \left[T_{s,k}\left(\mathbf{\Sigma}_{s,k}\right)^{-1} + \left(\mathbf{V}_{s,k}\right)^{-1}\right]^{-1}, \\
\mathbf{d}_{s,k} &= \left(\mathbf{\Sigma}_{s,k}\right)^{-1}\sum_{t=1}^{T}\left[\mathbf{Y}_t \cdot I\left(\tau_t = s, \lambda_t = k\right)\right] + \left(\mathbf{V}_{s,k}\right)^{-1}\mathbf{b}_{s,k}, \\
T_{s,k} &= \sum_{t=1}^{T} I\left(\tau_t = s, \lambda_t = k\right).
\end{aligned}$$

In other words, the posterior $\mu_{s,k}$ is determined by its prior as well as observations that fall into regime s and state k. Similarly, the posterior $\mathbf{\Sigma}_{s,k}$ is determined by its prior as well as observations that fall into the regime s and the state k. It follows that

$$\left(\mathbf{\Sigma}_{s,k}\right)^{-1}|\cdot \sim Whishart\left(\overline{\mathbf{\Omega}}_{s,k}, \overline{\nu}_{s,k}\right),$$

where

$$\overline{\mathbf{\Omega}}_{s,k} = \left\{ (\mathbf{\Omega}_{s,k})^{-1} + \sum_{t=1}^{T} \left[(\mathbf{Y}_t - \mu_{s,k}) (\mathbf{Y}_t - \mu_{s,k})' \cdot I(\tau_t = s, \lambda_t = k) \right] \right\}^{-1},$$

$$\overline{\nu}_{s,k} = \nu_{s,k} + T_{s,k}.$$

In models where $\mathbf{\Sigma}_{s,k}$ does not vary with the state k, the posterior $\mathbf{\Sigma}_{s,k}$ takes a similar form by replacing $I(\tau_t = s, \lambda_t = k)$ with $I(\tau_t = s)$ during the summation. If $\mathbf{\Sigma}_{s,k}$ further does not change with regimes, the summation is taken for the whole sample period.

The posterior of the mixture probability is given by

$$\eta_s \,|\cdot \sim Dirichlet\left(f_{s,1} + T_{s,1}, ..., f_{s,K} + T_{s,K}\right).$$

The posterior λ_t can take one of the K discrete states. The posterior distribution takes the form

$$P(\lambda_t = k \,|\cdot) \propto \prod_{s=1}^{S} \left[\eta_{s,k} \phi\left(\mathbf{Y}_t; \mu_{s,k}, \mathbf{\Sigma}_{s,k}\right)\right]^{I(\tau_t = s)},$$

where $k = 1, ...K$. It follows that $\lambda_t = k \,|\cdot$ has a multinomial distribution with probability proportional to $\prod_{s=1}^{S} \left[\eta_{s,k} \phi\left(\mathbf{Y}_t; \mu_{s,k}, \mathbf{\Sigma}_{s,k}\right)\right]^{I(\tau_t = s)}$.

The posterior of the initial distribution of the Markov chain is

$$\pi \,|\cdot \sim Dirichlet\left[c_1 + I(\tau_1 = 1), ..., c_S + I(\tau_1 = S)\right].$$

The posterior of the transition matrix of the Markov chain takes the form

$$\mathbf{Q}_s \,|\cdot \sim Dirichlet\left[a_{s,1} + \widetilde{T}_{s,1}, ..., a_{s,S} + \widetilde{T}_{s,S}\right],$$

where $\widetilde{T}_{s,j} = \sum_{t=1}^{T} I(\tau_{t-1} = s, \tau_t = j)$, $j = 1, ..., S$.

The posterior latent regimes τ_t can take one of the S discrete regimes. A straightforward method of sampling τ_t is to make use of its two neighbors τ_{t-1} and τ_{t+1}. However, MCMC chain may mix poorly due to excessive nodes on the chain. A better method is to sample the entire series τ_1^T by the Baum-Welch algorithm. The algorithm outlined here is similar to Chib (1996), who uses a backward induction. We sample the latent regimes in a forward sequence: τ_1, τ_2, τ_3, etc.

Let θ be all the parameters of the model (including $\mu_{s,k}$, $\mathbf{\Sigma}_{s,k}$, η_s, $\mathbf{Q}_s$, π).

Define the forward variable

$$F_{t,s} = P\left(\tau_t = s, \mathbf{Y}_1^t \left|\theta, \lambda_1^T\right.\right), t = 1, ..., T, s = 1, ..., S.$$

The forward variable can be computed by forward induction:

$$F_{t,s} = \left[\prod_{k=1}^{K} \phi\left(\mathbf{Y}_t; \mu_{s,k}, \mathbf{\Sigma}_{s,k}\right)^{I(\lambda_t = k)}\right] \cdot \sum_{r=1}^{S} F_{t-1,r} Q_{r,s},$$

Similarly define the backward variable

$$B_{t,s} = P\left(\mathbf{Y}_{t+1}^T \left|\theta, \lambda_1^T, \tau_t = s\right.\right),$$

which can be computed by backward induction:

$$B_{t,s} = \sum_{r=1}^{S} Q_{s,r} \cdot \left[\prod_{k=1}^{K} \phi\left(\mathbf{Y}_{t+1}; \mu_{r,k}, \mathbf{\Sigma}_{r,k}\right)^{I(\lambda_{t+1} = k)}\right] \cdot B_{t+1,r},$$

Note that $P\left(\tau_1^T \left|\mathbf{Y}_1^T, \theta, \lambda_1^T\right.\right) = P\left(\tau_1 \left|\mathbf{Y}_1^T, \theta, \lambda_1^T\right.\right) \cdot \prod_{t=2}^{T} P\left(\tau_t \left|\tau_{t-1}, \mathbf{Y}_1^T, \theta, \lambda_1^T\right.\right)$, so we sample τ_1^T by the method of composition.

For each $s, r = 1, ..., S$, we have

$$P\left(\tau_1 = s \left|\mathbf{Y}_1^T, \theta\right.\right) \propto F_{1,s} \cdot B_{1,s},$$

$$P\left(\tau_t = s \left|\tau_{t-1} = r, \mathbf{Y}_1^T, \theta\right.\right) \propto Q_{r,s} \cdot \left[\prod_{k=1}^{K} \phi\left(\mathbf{Y}_t; \mu_{s,k}, \mathbf{\Sigma}_{s,k}\right)^{I(\lambda_t = k)}\right] \cdot B_{t,s}.$$

The sampler for the latent regimes can be further improved by using the Gaussian mixture instead of normal distribution. Note that in the above procedure, τ_1^T is sampled from its full posterior conditionals, in which we explore the realization of the latent states λ_t in the mixture so that the term $\prod_{k=1}^{K}\phi\left(\mathbf{Y}_t;\mu_{s,k},\mathbf{\Sigma}_{s,k}\right)^{I(\lambda_t=k)}$ is effectively a normal density. However, if we put τ_1^T and λ_1^T in a block and sample them together using the method of composition, the nodes on the MCMC will be shortened and mixing property can be improved. In the block sampler, we first sample τ_1^T from it posterior distribution without being conditional on λ_1^T, then we sample λ_1^T from its full posterior conditionals. The previous procedure is modified by

$$\begin{aligned} F_{t,s} &= \left[\sum_{k=1}^{K}\eta_{s,k}\phi\left(\mathbf{Y}_t;\mu_{s,k},\mathbf{\Sigma}_{s,k}\right)\right]\cdot\sum_{r=1}^{S}F_{t-1,r}Q_{r,s}, \\ B_{t,s} &= \sum_{r=1}^{S}Q_{s,r}\cdot\left[\sum_{k=1}^{K}\eta_{r,k}\phi\left(\mathbf{Y}_{t+1};\mu_{r,k},\mathbf{\Sigma}_{r,k}\right)\right]\cdot B_{t+1,r}, \end{aligned}$$

and

$$P\left(\tau_t = s \left| \tau_{t-1} = r, \mathbf{Y}_1^T, \theta\right.\right) \propto Q_{r,s}\cdot\left[\sum_{k=1}^{K}\eta_{s,k}\phi\left(\mathbf{Y}_t;\mu_{s,k},\mathbf{\Sigma}_{s,k}\right)\right]\cdot B_{t,s}.$$

There is a note to the above Gibbs sampler. The hidden Markov models and Gaussian mixture models have an identification problem, that is, the likelihood function is invariant to regime/state label switching. There are some controversies over the interpretation of the label switching problem. Celeux et al. (2000) argue that virtually the entirety of MCMC samplers to the mixture model fails to converge due to the computational and inferential difficulties. Jasra et al. (2005) pessimistically believe that Gibbs sampler is not always appropriate for the mixture model. On the other side of the battle, Fruhwirth-Schnatter (2001) addresses the problem directly by adding a parameter random permutation step after each iteration of the simulator. Geweke (2007) insightfully points out that Gibbs sampler can reliably recover the posterior as long as the function of interest is invariant to permutation. He also proposes a conceptual permutation-augmented posterior simulator. Our function of interest is the posterior predictive asset returns whose distribution does not depend on the regime/state label. So the labeling phenomenon is not a problem.

4. Investor's Problem

Once we have a probability model on the asset returns, we are ready to solve an investor's portfolio optimization problem. It is most natural to assume the goal of portfolio selection is to maximize expected utility on future portfolio returns, though there are other ways to define the goal of portfolio optimization. For example, Buckley et al. (2008) consider maximizing the portfolio Sharpe ratio and out-performance probability of return target. When the asset returns are normally distributed, these goals are closed related and consistent with each other. However, different goals lead to different portfolios when asset returns depart from normality. In this section, we discuss an investor's problem in the expected utility maximization framework.

Let $u(\cdot)$ be a standard utility function. Assume the investor maximizes expected returns $E\left[u\left(\omega'\mathbf{Y}_{T+1}\right)\middle|\mathbf{Y}_1^T\right]$ subject to $\omega'\iota = 1$, where ω is the portfolio weights and ι is a vector of ones.[3]

[3]This maximization problem also depends on an investor's initial wealth. For example,

With MCMC we obtain simulated posterior sample of $\left\{\theta^{(j)}, \tau_1^{T,(j)}, \lambda_1^{T,(j)}\right\}_{j=1}^{J}$, where J is the number of draws in the simulation. The fact that

$$P\left(\mathbf{Y}_{T+1}, \tau_{T+1}, \lambda_{T+1}, \theta, \tau_T \left|\mathbf{Y}_1^T\right.\right)$$
$$= P\left(\theta, \tau_T \left|\mathbf{Y}_1^T\right.\right) \cdot P\left(\tau_{T+1} \left|\theta, \tau_T\right.\right) \cdot P\left(\lambda_{T+1} \left|\tau_{T+1}, \theta\right.\right) \cdot P\left(\mathbf{Y}_{T+1} \left|\lambda_{T+1}, \tau_{T+1}, \theta\right.\right)$$

suggests the following procedure of sampling $\mathbf{Y}_{T+1}$ from its posterior predictive distribution. First, sample the period $T+1$ latent regime $\tau_{T+1}^{(j)}$ using the information $\tau_T^{(j)}, \theta^{(j)}$. Second, sample the latent state $\lambda_{T+1}^{(j)}$ using $\tau_{T+1}^{(j)}, \theta^{(j)}$. Third, sample the asset prices $\mathbf{Y}_{T+1}^{(j)}$ using $\lambda_{T+1}^{(j)}, \tau_{T+1}^{(j)}, \theta^{(j)}$. It follows that $E\left[u\left(\omega'\mathbf{Y}_{T+1}\right)\left|\mathbf{Y}_1^T\right.\right]$ can be approximated by $\frac{1}{J}\sum_{j=1}^{J} u\left(\omega'\mathbf{Y}_{T+1}^{(j)}\right)$.

Note that solving an investor's problem requires choosing a portfolio weight ω to maximize the expected utility by some numerical optimization method. It poses a computational challenge in that numerical optimization is intermingled with simulation. If we want to run a large scale simulation (large J) and have many assets (large n), the computation cost might not be affordable. In that case, we may consider an approximation method that replaces expected utility with posterior moments in the optimization problem.

I borrow one dollar from my mom and repay the principal at the end of the period. My utility is defined on $\omega'\mathbf{R}_{T+1}$. Similarly, another investor earns a salary income of w_0 and then invest his one dollar in the stock market. His utility is defined on $w_0 + \omega'\mathbf{R}_{T+1}$. In this paper, we arbitrarily set $w_0 = 0$ in the investor's problem.

Define the moments of posterior predictive returns as

$$\begin{aligned}
\mathbf{M}_1 &= E\left(\mathbf{Y}_{T+1} \left| \mathbf{Y}_1^T \right.\right), \\
\mathbf{M}_2 &= E\left[\left(\mathbf{Y}_{T+1} - M_1\right)\left(\mathbf{Y}_{T+1} - M_1\right)' \left| \mathbf{Y}_1^T \right.\right], \\
\mathbf{M}_3 &= E\left[\left(\mathbf{Y}_{T+1} - M_1\right)\left(\mathbf{Y}_{T+1} - M_1\right)' \otimes \left(\mathbf{Y}_{T+1} - M_1\right)' \left| \mathbf{Y}_1^T \right.\right], \\
\mathbf{M}_4 &= E\left[\left(\mathbf{Y}_{T+1} - M_1\right)\left(\mathbf{Y}_{T+1} - M_1\right)' \otimes \left(\mathbf{Y}_{T+1} - M_1\right)' \otimes \left(\mathbf{Y}_{T+1} - M_1\right)' \left| \mathbf{Y}_1^T \right.\right].
\end{aligned}$$

The Kronecker product $\otimes$ enables us to spread the high-dimension array into a two-dimension matrix. Then the moments of the portfolio return can be expressed as

$$\begin{aligned}
M_{1p} &\equiv E\left(\omega'\mathbf{Y}_{T+1} \left| \mathbf{Y}_1^T \right.\right) = \omega'\mathbf{M}_1, \\
M_{2p} &\equiv E\left[\left(\omega'\mathbf{Y}_{T+1} - M_{1p}\right)^2 \left| \mathbf{Y}_1^T \right.\right] = \omega'\mathbf{M}_2\omega, \\
M_{3p} &\equiv E\left[\left(\omega'\mathbf{Y}_{T+1} - M_{1p}\right)^3 \left| \mathbf{Y}_1^T \right.\right] = \omega'\mathbf{M}_3\left(\omega \otimes \omega\right), \\
M_{4p} &\equiv E\left[\left(\omega'\mathbf{Y}_{T+1} - M_{1p}\right)^4 \left| \mathbf{Y}_1^T \right.\right] = \omega'\mathbf{M}_4\left(\omega \otimes \omega \otimes \omega\right).
\end{aligned}$$

Lastly, by a Taylor expansion of $E\left[u\left(\omega'\mathbf{Y}_{T+1}\right) \left| \mathbf{Y}_1^T \right.\right]$ up to order four, which accommodates the effects of mean, variance, skewness and kurtosis, we have

$$E\left[u\left(\omega'\mathbf{Y}_{T+1}\right) \left| \mathbf{Y}_1^T \right.\right] \approx u\left(M_{1p}\right) + \frac{1}{2}u''\left(M_{1p}\right)M_{2p} + \frac{1}{6}u^{(3)}\left(M_{1p}\right)M_{3p} + \frac{1}{6}u^{(4)}\left(M_{1p}\right)M_{4p}.$$

In practice, we only have posterior draws $\left\{\mathbf{Y}_{T+1}^{(j)}\right\}_{j=1}^{J}$, so the population moments are approximated by their sample analogues. For example,

$$\begin{aligned}
\widehat{\mathbf{M}}_1 &= \frac{1}{J}\sum_{j=1}^{J}\mathbf{Y}_{T+1}^{(j)}, \\
\widehat{\mathbf{M}}_3 &= \frac{1}{J}\sum_{j=1}^{J}\left(\mathbf{Y}_{T+1}^{(j)} - \widehat{M}_1\right)\left(\mathbf{Y}_{T+1}^{(j)} - \widehat{M}_1\right)' \otimes \left(\mathbf{Y}_{T+1}^{(j)} - \widehat{M}_1\right)',
\end{aligned}$$

and $\widehat{\mathbf{M}}_2$, $\widehat{\mathbf{M}}_4$ can be computed similarly. The analogue $\widehat{M}_{1p}, \widehat{M}_{2p}, \widehat{M}_{3p}, \widehat{M}_{4p}$ are computed from $\widehat{\mathbf{M}}_1, \widehat{\mathbf{M}}_2, \widehat{\mathbf{M}}_3, \widehat{\mathbf{M}}_4$, and the analogue moments of the portfolio return are used to approximate $E\left[u\left(\omega'\mathbf{R}_{T+1}\right)\middle|\mathbf{Y}_1^T\right]$. Note that it is not a "certainty equivalent solution". If the analogue moments are computed from data, the magnitude of estimation risk is fixed since it is impossible to increase the number of observations in the dataset. However, in our model the analogue moments are computed from posterior draws of future returns, the magnitude of estimation risk can be arbitrarily close to zero as long as we take large enough draws in the MCMC.

Using posterior moments to approximate expected utility reduces computational cost in that simulation is disentangled from numerical optimization. The analogue moments $\widehat{\mathbf{M}}_1, \widehat{\mathbf{M}}_2, \widehat{\mathbf{M}}_3, \widehat{\mathbf{M}}_4$ are computed with simulation before numerical optimization. In the stage of numerical optimization, only $\widehat{M}_{1p}, \widehat{M}_{2p}, \widehat{M}_{3p}, \widehat{M}_{4p}$ needs to be computed for each ω, which involves no simulation.

5. An application

To illustrate our approach, we consider a portfolio manager who diversifies investments in six world major stock indexes: SP500 (USA), FTSE (Britain), CAC (France), DAX (Germany), HSI (Hong Kong), NIKKEI 225 (Japan). Daily data ranging Jan. 2000-Dec. 2011 are used to estimate the asset returns.

Table 1 and 2 provides descriptive statistics of our dataset. Sample moments are calculated for entire sample from 2000 to 2011. The mean of index returns for most markets are negative, largely due to the global recession since 2008. In hindsight, it would be better off to lock the money in the coffer, rather than to invest any dollar in the stock market. For illustration purposes, we exclude the possibility of refraining from investment and assume no safe assets. The covariance matrix of returns suggests stronger positive correlations among western countries. The correlations between western and oriental markets are less prominent, which carries significance for global diversifications.

As is seen in Table 1, daily returns of stock indexes exhibit substantial skewness for most markets. 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.

The departure from normality can also be seen from the Bayesian residual test. We first fit the MSGM model with one regime and one state, which is effectively a model of multivariate normal returns. We conduct a series of residual tests by normalizing the historical returns using the posterior draws of the mean and covariance matrix. If the returns are indeed normally distributed, then the classical Kolmogorov-Smirnov test should accept the null. The histogram of the test statistics are reported in Figure 1. The six panels correspond to the six assets in sequence. Since we have a fairly large sample size of more than 2000 observations, the 1% significance critical value of the test statistics can be approximated by $1.63/\sqrt{T}$, which is about 0.03. Figure 1 shows that test statistics are larger than the critical value in every circumstance so that the normality can be decisively rejected.

We also go through the above tests for subsamples of the data set. Without an exception, the normality of asset returns is rejected. For brevity, we did not report the results in the text. Noted that all the normality tests are conducted on the basis of individual asset returns. Once the normality is rejected at individual level, the joint normality is automatically rejected (The reverse is not true). We therefore conclude that it is necessary to adopt more flexible distributions to model asset returns.

In this application, we fit the MSGM model with three Markov switching regimes, each with three states in the Gaussian mixture. The covariance matrix is assumed to be invariant across regimes and states. The posterior predictive distributions of asset returns are reported in Table 3. The predicted mean returns are positive, in contrast to the slightly negative mean returns over the entire sample. The positive returns prediction might due to the fact that at the end of our sample period, the return series tend to be positive and thus in a high-return regime. The variances of the predictive distributions contain triple uncertainties, namely the uncertainty over the regimes and states, the uncertainty over the parameters and the uncertainty over the future disturbances. However, Table 3 shows that for each asset the predictive variance is smaller than the sample variance (which effectively corresponds to a model with one regime and one state). That implies the MSGM model better captures the non-normality feature of the data and improves precision of the prediction.

Using the draws from the posterior predictive distribution, Figure 2 plots the Bayesian mean-variance frontier. For comparison, we also provide the mean-variance frontier with certainty equivalent approach. The two curves present different mean-variance trade-offs. For a given variance, the Bayesian method predicts a higher expected return.

Mean-variance frontier may not be directly relevant to decision making in the presence of non-normality. The next step is to estimate the optimal portfolio weights which maximize the expected utility. We use a Tylor approximation up to the fourth order and assume that the portfolio manager has a CARA utility. Table 4 shows the optimal weights with risk aversion coefficients 1, 3, 5, 7, 9 and 11. The optimal weight on the third and fourth assets are negative, which requires short selling. Table 5 provides optimal weights when short selling is not allowed.

6. Conclusion

Departure from normality and parameter estimation risk are two major barriers to the implementation of the Markowitz portfolio selection. This paper attempts to addresses the two issues in a unified Bayesian framework, in which deviation from normality is captured by a Markov switching Gaussian mixture distribution and parameter uncertainty is reflected in the posterior predictive distribution of asset returns. 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 of investing in several leading stock market indexes. The descriptive statistics provide strong evidence against normality of high frequency index returns. A model with four regimes and four states is used to predict the future returns, and the associated optimal portfolios are also reasonably diversified among assets.

Ane, T., Geman, H., 2000. Order flow, transaction clock, and normality of asset returns. Journal of Finance 55 (5), 2259–2284.

Baum, L. E., Eagon, J. A., 1967. An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology. Bulletin of American Mathematical Society 73, 360–363.

Baum, L. E., Petrie, T., 1966. Statistical inference for probabilistic functions of finite state markov chains. The Annals of Mathematical Statistics 37 (6), 1554–1563.

Baum, L. E., Petrie, T., Soules, G., Weiss, N., 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. The Annals of Mathematical Statistics 41 (1), 164–171.

Black, F., Litterman, R., 1992. Global portfolio optimization. Financial Analysts Journal 48, 28C43.

Blattberg, R. C., Gonedes, N. J., 1974. A comparison of the stable and student distributions as statistical models for stock prices. The Journal of Business 47 (2), 244–80.

Buckley, I., Saunders, D., Seco, L., 2008. Portfolio optimization when asset returns have the gaussian mixture distribution. European Journal of Operational Research 185 (3), 1434–1461.

Calvet, L. E., Fisher, A. J., 2004. How to forecast long-run volatility: Regime switching and the estimation of multifractal processes. Journal of Financial Econometrics 2 (1), 49–83.

Celeux, G., Hurn, M., Robert, C. P., 2000. Computational and inferential difficulties with mixture posterior distributions. Journal of the American Statistical Association 95 (451), 957–970.

Chib, S., 1996. Calculating posterior distributions and modal estimates in markov mixture models. Journal of Econometrics 75 (1), 79–97.

Chopra, V. K., Ziemba, W. T., 1993. The effect of errors in means, variances, and covariances on optimal portfolio choice. Journal of Portfolio Management 19, 6–11.

Dickinson, J. P., 1974. The reliability of estimation procedures in portfolio analysis. Journal of Financial and Quantitative Analysis 9 (03), 447–462.

Fama, E. F., 1965. The behavior of stock-market prices. The Journal of Business 38 (1), pp. 34–105.

Fruhwirth-Schnatter, S., 2001. Markov chain monte carlo estimation of classical and dynamic switching and mixture models. Journal of the American Statistical Association 96 (453), 194–209.

Geweke, J., 2007. Interpretation and inference in mixture models: Simple mcmc works. Computational Statistics and Data Analysis 51 (7), 3529–3550.

Hamilton, J., 2005. Regime-switching models. dss.ucsd.edu/ jhamilto/palgrav1.pdf.

Hamilton, J. D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57 (2), 357–384.

Harvey, C., Liechty, J., Liechty, M., Muller, P., 2010. Portfolio selection with higher moments. Quantitative Finance 10 (5), 469–485.

Harvey, C. R., Siddique, A., 2000. Conditional skewness in asset pricing tests. Journal of Finance 55 (3), 1263–1295.

Jasra, A., Holmes, C. C., Stephens, D. A., 2005. Markov chain monte carlo methods and the label switching problem in bayesian mixture modeling. Statistical Science 20 (1), 50–67.

Jobson, J. D., Korkie, B., 1980. Estimation for markowitz efficient portfolios. Journal of the American Statistical Association 75 (371), 544–554.

Jondeau, E., Rockinger, M., 2006. Optimal portfolio allocation under higher moments. European Financial Management 12 (1), 29–55.

Kim, C.-J., Piger, J., Startz, R., 2008. Estimation of markov regime-switching regression models with endogenous switching. Journal of Econometrics 143 (2), 263–273.

Kraus, A., Litzenberger, R. H., 1976. Skewness preference and the valuation of risk assets. Journal of Finance 31 (4), 1085–1100.

Krolzig, H.-M., 1997. Markov-switching vector autoregressions : modelling, statistical inference, and application to business cycle analysis. Springer.

Lucas, R. E., 1978. Asset prices in an exchange economy. Econometrica 46 (6), 1429–1445.

Markowitz, H., 1952. Portfolio selection. Journal of Finance 7, 77–91.

Michaud, R. O., Michaud, R. O., 2008. Efficient asset management: A practical guide to stock portfolio optimization and asset allocation.

Peiro, A., 1999. Skewness in financial returns. Journal of Banking and Finance 23 (6), 847 – 862.

Rabiner, L. R., 1989. A tutorial on hidden markov models and selected applications in speech recognition. In: Proceedings of the IEEE. pp. 257–286.

Roman, D., Mitra, G., Spagnolo, N., 2010. Hidden markov models for financial optimization problems. IMA Journal of Management Mathematics 21 (2), 111–129.

Scherer, B., 2002. Portfolio resampling: Review and critique. Financial Analysts Journal 58 (6), 98–109.

Sims, C. A., Zha, T., 2006. Were there regime switches in u.s. monetary policy? American Economic Review 96 (1), 54–81.

Turner, C. M., Startz, R., Nelson, C. R., 1989. A markov model of heteroskedasticity, risk, and learning in the stock market. Journal of Financial Economics 25 (1), 3–22.

	SP500	FTSE	CAC	DAX	HSI	NIKKEI
Mean	-0.005	-0.008	-0.024	-0.005	0.003	-0.031
Skewness	-0.193	0.013	0.117	0.010	-0.216	-0.408
Kurtosis	9.180	9.818	8.667	8.585	12.586	9.528
Lilliefors	0.080	0.079	0.071	0.068	0.080	0.058
p-val	0.001	0.001	0.001	0.001	0.001	0.001
Jarque-Bera	4258.0	5162.4	3572.5	3463.4	10225.5	4806.2
p-val	0.001	0.001	0.001	0.001	0.001	0.001

Table 1: Descriptive statistics of daily percentage asset returns

	SP500	FTSE	CAC	DAX	HSI	NIKKEI
SP500	2.123	1.191	1.500	1.659	0.705	0.490
FTSE	1.191	1.966	2.150	2.016	1.135	0.935
CAC	1.500	2.150	2.907	2.675	1.362	1.150
DAX	1.659	2.016	2.675	3.118	1.321	1.066
HSI	0.705	1.135	1.362	1.321	3.109	1.911
NIKKEI	0.490	0.935	1.150	1.066	1.911	2.811

Table 2: Sample covariance matrix of daily percentage asset returns

	Pred. Mean	Pred. Var	$\Pr(Y > 0)$
SP500	0.020	1.567	0.505
FTSE	0.034	1.407	0.511
CAC	0.016	2.246	0.506
DAX	0.032	2.342	0.512
HSI	0.071	2.800	0.518
NIKKEI	0.036	2.621	0.509

Table 3: Summary of the posterior predictive asset returns with a three-regime, three-state MSGM model

	1	3	5	7	9	11
SP500	40.5	39.5	38.3	37.1	35.9	34.7
FTSE	83.3	87.8	90.2	92.4	94.7	97.0
CAC	-33.5	-30.0	-31.5	-33.5	-35.5	-37.7
DAX	-17.0	-22.5	-22.4	-21.9	-21.3	-20.6
HSI	6.9	1.2	0.8	1.3	2.0	2.9
NIKKEI	19.7	24.1	24.7	24.6	24.2	23.8

Table 4: Optimal portfolio weights (in percentage) under different risk aversion coefficients while short selling is allowed. The expected utility is approximated by the Taylor expansion of order four.

	1	3	5	7	9	11
SP500	33.4	29.3	26.6	26.3	25.6	24.5
FTSE	44.9	52.7	54.8	55.7	56.2	55.9
CAC	0.0	0.0	0.0	0.0	0.0	0.0
DAX	0.0	0.0	0.9	0.0	0.1	1.1
HSI	4.6	2.2	0.6	0.2	0.0	1.2
NIKKEI	17.1	15.8	17.1	17.8	18.2	17.3

Table 5: Optimal portfolio weights (in percentage) under different risk aversion coefficients while short selling is not allowed. The expected utility is approximated by the Taylor expansion of order four.

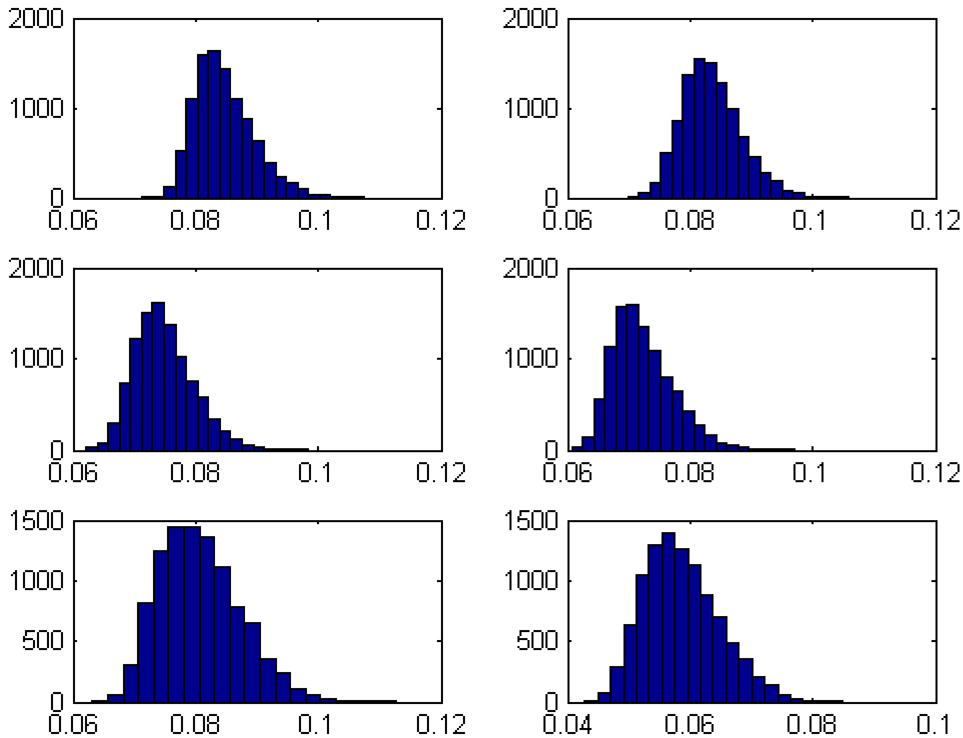

Figure 1: Bayesian residual Kolmogorov-Smirnov test statistics. The six panels correspond to SP500, FTSE, CAC, DAX, HSI, NIKKEI respectively. Under the null of normality, the critical value of 1% significance level is approximately 0.03.

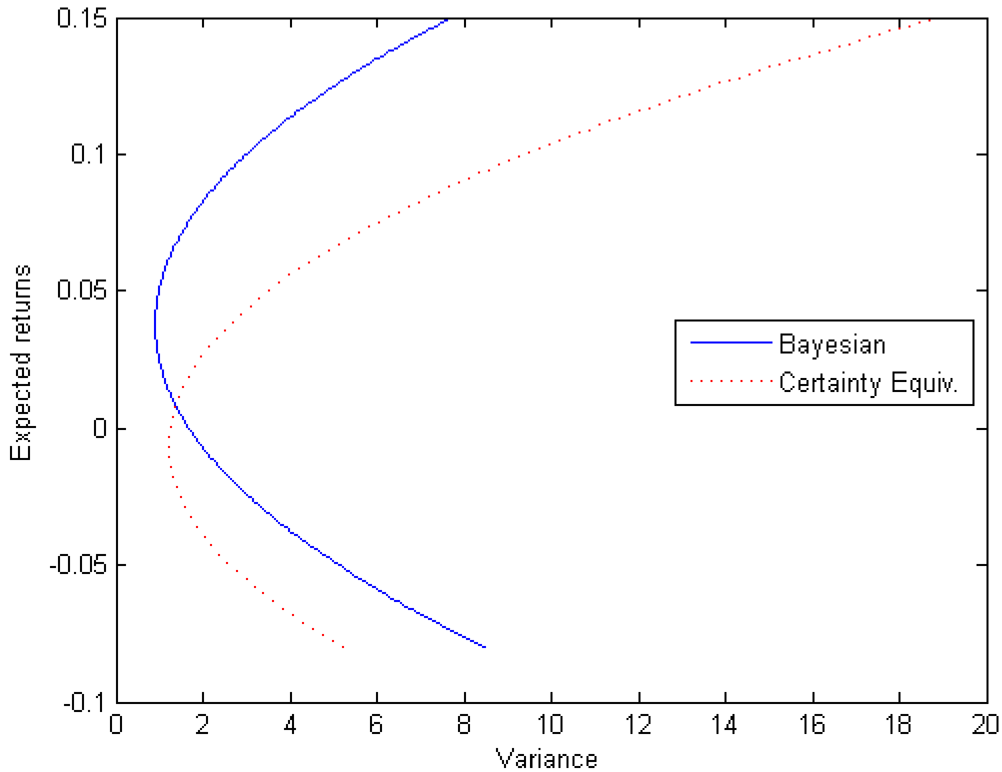

Figure 2: A comparison of the classic mean variance frontier (certainty equivalence solution) with the mean variance frontier using the posterior predictive distribution of asset returns with the MSGM model.