Available online at www.sciencedirect.com
ScienceDirect
Economic Modelling 24 (2007) 329 – 349
Economic Modelling
www.elsevier.com/locate/econbase

# Estimation of a utility-based asset pricing model using normal mixture GARCH(1,1)

C.C. Wu [a], Jack C. Lee [b,*]

[a] Department of Management Science, National Chiao Tung University, Hsinchu, Taiwan
[b] Graduate Institute of Finance and Institute of Statistics, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu, 300 Taiwan

Accepted 15 August 2006

**Abstract**

Brown and Gibbons [Brown, D.P., Gibbons, M.R., 1985. A simple econometric approach for utility-based asset pricing model. Journal of Finance 40, 359–381], Karson et al. [Karson, M., Cheng, D., Lee, C. F., 1995. Sampling distribution of the relative risk aversion estimator: theory and applications. Review of Quantitative Finance and Accounting 5, 43–54], and Lee et al. [Lee, C.F., Lee, J.C., Ni, H.F., Wu, C.C., 2004. On a simple econometric approach for utility-based asset pricing model. Review of Quantitative Finance and Accounting 22, 331–344] developed the theory and the distribution of unconditional relative risk aversion (RRA) estimates in utility-based asset pricing model by assuming normality for the log excess returns. While the normality assumption is not always appropriate for some security returns, Brown and Gibbons [Brown, D.P., Gibbons, M.R., 1985. A simple econometric approach for utility-based asset pricing model. Journal of Finance 40, 359–381] proposed generalized method of moments (GMM) to estimate unconditional RRA. However, RRA estimated by GMM is not statistically efficient with finite samples. The main purpose of this paper is to derive the process of estimating dynamic RRA with the maximum likelihood and a Bayesian method having a weakly informative prior density while assuming that the log excess returns on the market are distributed as normal mixture GARCH(1,1). This methodology will capture the variations of RRA across different periods. Empirical results are presented using market rates of returns and risk-free rates over the period 1941 to 2001.


*JEL classification:* C11; C15; D31

*Keywords:* Relative risk aversion; MLE; Bayesian

* Corresponding author. Tel.: +886 3 5131568; fax: +886 3 5733260.
*E-mail address:* jclee@stat.nctu.edu.tw (J.C. Lee).

0264-9993/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.econmod.2006.08.003

## 1. Introduction

For the last several decades, utility-based asset pricing models have been investigated extensively and attracted some empirical attention in financial economics. The power utility function defined over consumption states with coefficient of relative risk aversion (RRA) and rate of time preference is the most commonly utilized preference specifications. Hansen and Singleton (1982, 1983) postulated that the pricing kernel is a power function of aggregate U.S. consumption and estimated the parameters of the power pricing kernel. Hansen and Jagannathan (1991) derived bounds of the consumption-based pricing kernel from the mean and standard deviation of the market portfolio excess returns.

Furthermore, because the unknown parameter of isoelastic utility function is RRA, Brown and Gibbons (1985) and Alonso et al. (1990) have argued that using this utility function can provide precious information for several reasons. First, since some theoretical results in finance rely on log utility function (i.e., $\beta=1$)(see, Hakanson (1970), Kraus and Litzenberger (1975), Rubinstein (1977), Cox, Ingersoll, and Ross (1985)), we must make appropriate judgment on these results when RRA is significantly different from one. Second, when we are confronted with the demand for risky assets and the savings decisions, the demand for risky assets depends on the magnitude of RRA (see, Rothschild and Stiglitz, 1971). Third, there are many research papers dealing with the issue whether stock prices have excessive volatility relative to the degree of aggregate risk aversion (see, Grossman and Shiller, 1981).

Moreover, Lucas (1978), Grossman and Shiller (1981), Duan and Singleton (1986) used aggregate consumption data for empirical analyses. However, Ermini (1989), Wilcox (1992), and Slesnick (1998) have discussed that analyses in terms of aggregate consumption data are affected by measurement problems, such as coding errors, imputation procedures, definitional problems, and sampling error. In addition, Campbell (1993) and Rosenberg and Engle (2002) have indicated that studies in use of these data, which are measured with error and are time-aggregated, will have serious consequence for asset pricing relationships. In order to avoid these problems, Brown and Gibbon (1985), Bansal and Viswanatham (1993), Campbell (1993), and Rosenberg and Engle (2002) replaced the aggregate consumption return with a proxy for the market portfolio return if the consumption was a constant proportion of wealth. Rosenberg and Engle (2002) also pointed out that pricing model estimation over equity return states may yield more intuition into investor risk aversion than estimation over National Income and Product Accounts (NIPA) consumption states if the market return is a better proxy of the true consumption return than the data from NIPA.

On the other hand, analyzing the distribution of asset returns data has been an important research area in financial economics. Accordingly, utility-based models of the asset pricing are of particular interest while the distribution of returns can be suitably determined and explained. When the excess return on the market portfolio is distributed as a lognormal distribution, Brown and Gibbons (1985), Karson et al. (1995), and Lee et al. (2004) have proposed different methods for estimating unconditional relative risk aversion (RRA), $\beta$, and derived the exact sampling and Bayesian estimators of RRA. However, it is well known that the normal distribution may not be adequate for the log asset returns. The empirical findings seem to indicate that the unconditional and conditional distributions of log asset returns are not symmetric and have fat tails relative to the normal distribution.

Therefore, Brown and Gibbons (1985) dropped the distributional assumption and recommended using generalized method of moments (GMM) to estimate unconditional RRA. GMM estimates and their standard errors are consistent even though residuals are heteroskedastic. However, the GMM can be applied only to large samples. In most cases GMM estimates are asymptotically efficient, but they are hardly efficient at finite samples. In addition, we found that the unconditional RRA

estimates vary a lot across different sub-periods in Brown and Gibbons (1985) and, Lee et al. (2004). Hence, estimating RRA based on unconditional and fixed moments will not capture the phenomenon of structural changes.

Even though a variety of evaluation methods have been proposed and implemented to RRA estimator to date, they have mostly depended on GMM or the classical statistical framework. However, parameter estimation risk and model mispricing risk are usually ignored in the classical statistical framework. Therefore, Pastor and Stambaugh (2000) assumed normality of returns and presented the Bayesian set-up, which factored uncertainties in both parameter estimation and model mispricing into investors' decision making.

Although many appropriate distributions have been proposed to analyze asset returns, the RRA estimator does not always have a solution by assuming asset returns distributed as any distribution. Thus, this paper extended Brown and Gibbons (1985) and used the NM($K$)-GARCH model, i.e. the model where errors have $K$-component normal mixture distribution with generalized autoregressive conditional heteroscedasticity (GARCH) variance process, to obtain more efficient dynamic RRA estimator and its distribution. The primary purpose of this paper is to derive the process of estimating dynamic RRA while assuming that the log excess returns on the market are distributed as NM($K$)-GARCH(1,1). On the part of the parameter estimation, we not only present the classical maximum likelihood but also recommend a Bayesian method, which combines an investor's prior belief about the accuracy of the pricing model and the information in the data, with a weakly informative prior density.

The rest of this paper is organized as follows. In Section 2, we briefly review the literature of RRA estimation. In Section 3, we present the NM($K$)-GARCH(1,1) model, maximum likelihood estimate of RRA, and a Bayesian approach to estimate RRA. Some simulation studies are shown in Section 4. In Section 5, we describe the data and RRA estimation results using three different approaches. Finally, we conclude the results in Section 6.

## 2. Literature review

Brown and Gibbons (1985) used the well known Euler condition for the dynamic consumption-portfolio problem faced by a representative individual under uncertainty to derive a relationship between RRA and the moments of security returns. In their work, the utility function in period $t$ from consumption $C_t$ follows power (or isoelastic) utility, that is,

$$U(C_t) = \frac{C_t^{1-\beta_t} - 1}{1-\beta_t} \tag{1}$$

where $\beta_t = -U''(C_t)C_t/U'(C_t)$ is RRA at time $t$. Furthermore, in order to avoid measurement problems with consumption, most early researches on utility-based asset pricing models replaced aggregate consumption with the return on some proxy for the market portfolio. Through above assumptions, the first-order necessary condition for optimality with a time additive derived by Brown and Gibbons (1985) is

$$E[(x_t - 1) \cdot x_t^{-\beta_t} | \Im_{t-1}] = 0 \tag{2}$$

where

| | |
|---|---|
| $\Im_{t-1}$ | Information set available to the market in period $t-1$, |
| $x_t$ | $(1+R_{mt})$ $(1+R_{ft})$, the excess return on the market, |
| $R_{mt}$ | Return on the market portfolio, |
| $R_{ft}$ | Return on the riskless asset. |

Assuming the existence of relevant moments, the law of iterated expectations applied to Eq. (2) implies

$$E[(x_t-1)x_t^{-\beta_t}] = 0. \tag{3}$$

Although individuals have changing conditional expectations through time, the econometrician, by relying on Eq. (3), can still construct a valid test of the theory based on unconditional and fixed moments. Hence, most previous literatures claimed that it is not necessary to specify a model for the conditional expectations or even the variables which affect these conditional expectations and only estimated the static RRA, called $\beta$ instead of $\beta_t$ in this paper, during some specific periods.

Rubinstein (1976) proposed estimating $\beta$ by assuming that the excess return on the market has a lognormal distribution, i.e., $\log x_t \sim N(\mu,\sigma^2)$, and derived

$$\beta = \frac{\mu}{\sigma^2} + \frac{1}{2}. \tag{4}$$

Following Brown and Gibbons (1985), a natural maximum likelihood estimator for $\beta$ is

$$\hat{\beta} = \frac{\hat{\mu}}{\hat{\sigma}^2} + \frac{1}{2},$$

where $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} \log x_t$ and $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(\log x_t - \overline{\log x_t})^2$.

Using the asymptotic theory, they also derived the variance of $\sqrt{T}\ \hat{\beta}$ as:

$$\mathrm{Var}(\sqrt{T}\ \hat{\beta}) = \frac{2[E(\log x_t)]^2 + \mathrm{Var}[\log x_t]}{[\mathrm{Var}(\log x_t)]^2}. \tag{5}$$

Alternatively, following Karson et al. (1995), the minimum variance unbiased (MVU) estimator of $\beta$ is

$$\hat{\beta} = \frac{(T-3)\ \hat{\mu}}{(T-1)\ \hat{\sigma}^2} + \frac{1}{2}. \tag{6}$$

Brown and Gibbons (1985) pointed out that estimate of $\beta$ is inconsistent when the distribution of log excess returns is a departure from normality. In order to remedy this weakness, they made use of the generalized method of moments to estimate RRA. Therefore, RRA could be estimated by finding the value $\hat{\beta}$ which satisfies the equation,

$$f(\ \hat{\beta}) = \frac{1}{T}\sum_{t}(x_t-1)x_t^{-\ \hat{\beta}} = 0, \tag{7}$$

where $\hat{\beta} \in (0,+\infty)$ , and the asymptotic variance is

$$\mathrm{Var}(\sqrt{T}\ \hat{\beta}) = \frac{E\{[(x_t-1)x_t^{-\ \hat{\beta}}]^2\}}{[E\{(x_t-1)x_t^{-\beta}\log x_t\}]^2}. \tag{8}$$

## 3. Methodology and estimation

The finite mixture model has a long and illustrious history in statistics. Starting with Newcomb's (1886) application of the normal mixture model for outliers, it has provided a mathematics-based approach to the statistical modeling of a wide variety of random phenomena. The maximum

likelihood estimation and its inference have become very popular with the advent of the EM algorithm of Dempster et al. (1977). Furthermore, with the advent of inexpensive, high-speed computers and the rapid development in posterior simulation techniques such as Markov chain Monte Carlo (MCMC) methods for enabling the Bayesian analysis to be undertaken, practitioners are increasingly turning to Bayesian methods for the analysis of complicated statistical models. Historical literatures, the latest inferential developments, and the applications associated with finite mixture models are given in Titterington, Smith, and Makov (1985), Mclachlan and Peel (2000).

Owing to its greater flexibility, the normal mixture model has also been found superior for describing the distribution of asset returns. Kon (1984) found positive skewness in the individual returns of the stocks that make up the Dow Jones Index and proposed a mixture of normal distributions as a suitable framework for capturing the skewness. Boothe and Glassman (1987) also argued that the normal mixture distribution offers a better description of exchange rate returns than the normal model.

The family of $K$-component normal mixtures is capable of exhibiting the skewness and excess kurtosis characteristics of economical and financial data. The advantages of normal mixtures are the valid distributional assumption and the ease in financial explanations. For example, the normal mixture formulation allows for a sensible interpretation of two or more heterogeneous groups of market participants. Consequently, bullish and bearish investors could behave differently. Moreover, the mixture can also be interpreted as representing trading days of different types. For instance, returns on the stock market on Mondays will follow the prevailing trend from the previous Friday, and this phenomenon is viewed as the Monday effect. Besides, there are also January effect, October effect, etc., and these phenomena may cause the stock's behavior in specific months to depart from the other trading months. Furthermore, some substantial political or economical policies may give rise to a component with relatively high or low variance and smaller weight.

Since Engle (1982) first found that volatility shows a clustering phenomenon, numerous academic literatures have proposed a variety of models to predict future volatilities and exhibit market volatilities which are more predictable than market returns. However, most of those papers assume that the conditional distributions of the asset returns are symmetric, such as normal or student-$t$ models. Recently, Haas, Mittnik and Paolella (2004) improved these drawbacks and proposed NM($K$)-GARCH(p,q) models with an inter-dependent autoregressive evolution for the variance series to capture the non-zero conditional excess kurtosis and sknewness. However, the results of Haas et al. (2004) exhibited that neither the dependence of component variances nor the inclusion of more than one lag in the conditional variance equations are significant. Therefore, Alexander and Lazar (2005) recommended using a simpler NM($K$)-GARCH(1,1) models to exchange rate modeling. Throughout this paper, we estimate RRA by applying NM($K$)-GARCH (1,1) model to log market excess returns.

The log market excess returns, $Y_t$, distributed as NM($K$)-GARCH(1,1), is specified as follows:

$$\begin{aligned} &Y_t|\Im_{t-1} \sim NM(p_1, \ldots, p_K; \mu_1, \ldots, \mu_K; \sigma_{1,t}^2, \ldots, \sigma_{K,t}^2), \\ &\sigma_{j,t}^2 = w_j + a_j\varepsilon_{t-1}^2 + b_j\sigma_{j,t-1}^2, \end{aligned} \tag{9}$$

where $\sum_{j=1}^{K} p_j = 1$, $\sum_{j=1}^{K} p_j\mu_j = \mu$, $\varepsilon_{t-1} = y_{t-1} - \mu$, $w_j > 0$, $a_j, b_j \geq 0$, and $a_j + b_j < 1$ for $j = 1, \ldots, K$. The conditional density function of $y_t$ is $f(y_t|\Im_{t-1}) = \sum_{j=1}^{K} p_j\phi_j(y_t|\Im_{t-1})$ and $\phi_j(y_t|\Im_{t-1})$ is normal density function with mean $\mu_j$ and variance $\sigma_{j,t}^2$.

For the mixture model, the number of required component densities is unknown and need to be empirically determined generally. Unfortunately, standard likelihood ratio test (LRT) theory breaks down in this framework (see, McLachlan and Peel, 2000). However, standard model

selection criteria such as AIC (Akaike, 1973) and BIC (Schwarz, 1978) are widely used to compare models with different numbers of components. For a model with $d$ parameters, sample size $T$ and log-likelihood $\ell$, evaluated at the maximum likelihood estimator, AIC $= -2\ell + 2d$ and BIC $= -2\ell + d \log T$. When $\log T > 2$, it can be found that the penalty term of BIC penalizes complex models more heavily than AIC. Hence, AIC tends to fit too many components, and BIC is more conservative in that it favors more parsimonious models. Furthermore, the literatures on mixtures provide some encouraging evidence in the context of unconditional model and suggest that BIC provides a reasonably good indication for the number of components (see, Roder and Wassermann, 1997; McLachlan and Peel, 2000). For this reason, we will use BIC to determine the number of required component densities in empirical studies.

### 3.1. ML estimation of dynamic RRA

Given a sample of $T$ observations, $y_t$, from the NM($K$)-GARCH(1,1) model, the conditional likelihood function is

$$L(\Psi) = \prod_{t=2}^{T} \left[ \sum_{j=1}^{K} p_j \phi_j(y_t | \Im_{t-1}) \right], \tag{10}$$

where $\Psi$ are all parameters of the NM($K$)-GARCH(1,1) model. Maximization of $L(\Psi)$ with respect to $\Psi$, for given data $\boldsymbol{y} = (y_1, y_2, \ldots, y_T)$, yields the maximum likelihood estimate of $\Psi$. Equivalently, and more usually, the quantity to be maximized is the log-likelihood $\ell(\Psi) = \log L(\Psi)$. For simple parametric models, he maximum likelihood approach is very popular, partly because it fits into the philosophy of likelihood-based inference (consistency, asymptotic efficiency, and invariance), partly because of the existence of attractive asymptotic normality theory, and partly because the estimates are often easy to compute. For the normal mixture model, however, we discover that the computational aspect is not always so straightforward. Therefore, we will apply the EM algorithm (see, Aitkin and Tunnicliffe Wilson, 1980), with the complete data conditional likelihood for $\Psi$ given by

$$L_c(\Psi) = \prod_{t=2}^{T} \left[ \sum_{j=1}^{K} p_j^{z_{tj}} [\phi_j(y_t | \Im_{t-1})]^{z_{tj}} \right], \tag{11}$$

where $z_{tj} = 1$ or 0 according to whether $y_t$ did or did not arise from the $j$th component of the mixture. Then, the complete data conditional log-likelihood is

$$\ell_c(\Psi) = \sum_{t=2}^{T} \sum_{j=1}^{K} z_{tj} \left[ \log p_j + \log \phi_j(y_t | \Im_{t-1}) \right]. \tag{12}$$

Suppose that $\Psi$ is known and equal to $\Psi^{(m)}$, the observed complete log-likelihood is given by

$$\begin{aligned} Q(\Psi; \Psi^{(m)}, y) &= E|\ell_c(\Psi)|y, \Psi^{(m)}| \\ &= \sum_{t=2}^{T} \sum_{j=1}^{K} \left[ E[z_{tj}|y, \Psi^{(m)}] \times [\log p_j + \log \phi_j(y_t | \Im_{t-1})] \right] \\ &= \sum_{t=2}^{T} \sum_{j=1}^{K} \left[ w_{tj}^{(m)}(\Psi^{(m)}) \times [\log p_j + \log \phi_j(y_t | \Im_{t-1})] \right] \end{aligned} \tag{13}$$

where $w_{tj}^{(m)}(\Psi^{(m)}) = p_j^{(m)} \phi_j(y_t | \Im_{t-1}, \Psi^{(m)}) \Big/ \sum_{j=1}^{K} p_j^{(m)} \phi_j(y_t | \Im_{t-1}, | \Psi^{(m)})$.

Therefore, the *EM Algorithm* is as follows:

1. E-step: Compute $Q(\Psi;\Psi^{(m)}, \boldsymbol{y})$.
2. M-step: Maximize numerical $Q(\Psi;\Psi^{(m)}, \boldsymbol{y})$ with respect to $\Psi$ and get updated estimates of the parameters, denoted by $\Psi^{(m+1)}$.
3. Repeat E-step and M-step until $\ell(\Psi)$ converging to a local maximum. We stop the iteration if $|\ell(\Psi^{(m+1)})-\ell(\Psi^{(m)})|<10^{-8}$.

For a transformed version of Eq. (2), we assume that log $x_t$, the log excess returns on the market, follow NM($K$)-GARCH(1,1) model. So,

$$y_t|\Im_{t-1} = \log x_t|\Im_{t-1} \sim \sim \mathrm{NM}(p_1, \ldots, p_K; \mu_1, \ldots, \mu_K; \sigma_{1,t}^2, \ldots, \sigma_{K,t}^2),$$

And consequently we have

$$E[x_t^{-\beta_t}|\Im_{t-1}] = \sum_{j=1}^{K} p_j e^{-\mu_j\beta_t+\frac{\sigma_{j,t}^2\beta_t^2}{2}}. \tag{14}$$

Substitution in Eq. (2) gives

$$g(\beta_t|\Im_{t-1}, \mu_j, \sigma_{j,t}^2, j=1, \ldots, K) = \sum_{j=1}^{K} p_j e^{\mu_j(1-\beta_t)+\frac{\sigma_{j,t}^2(1-\beta_t)^2}{2}} - \sum_{j=1}^{K} p_j e^{-\mu_j\beta_t+\frac{\sigma_{j,t}^2\beta_t^2}{2}} = 0. \tag{15}$$

Given that the excess returns on the market portfolio is NM($K$)-GARCH(1,1) model, Eq. (15) provides a relationship between dynamic RRA and the parameters of the NM($K$)-GARCH(1,1) model. Since log $x_t|Z_{t-1}$ has a normal mixture distribution, the MLE of RRA has no analytical form. However, because a continuous function of maximum likelihood estimators is also a maximum likelihood estimator (Zehna, 1966), $\hat{\beta}_t$ satisfying the following equation has all of the well known properties of the maximum likelihood estimator,

$$g(\hat{\beta}t|\Im_{t-1}, \hat{\mu}_j, \hat{\sigma}_{j,t}^2, j=1, \ldots, K) = \sum_{j=1}^{K} \hat{p}_j e^{\hat{\mu}_j(1-\hat{\beta}_t)+\frac{\hat{\sigma}_{j,t}^2(1-\hat{\beta}_t)^2}{2}} - \sum_{j=1}^{K} \hat{p}_j e^{-\hat{\mu}_j\hat{\beta}_t+\frac{\hat{\sigma}_{j,t}^2\hat{\beta}_t^2}{2}} = 0. \tag{16}$$

where $\hat{\sigma}_{j,t}^2 = \hat{w}+\hat{a}_j\varepsilon_{t-1}^2+\hat{b}_j\,\sigma_{j,t-1}^2$ for $j=1,..,K$.

In the Appendix, we prove that the function $g(\beta|\Im_{t-1}, \mu_j, \sigma_{j,t}^2\, j=1\ldots K)$ is a strictly decreasing continuous function, so Eq. (16) has a unique solution. In addition, Fig. 1 illustrates the nonlinear search for $\beta_t$, when the other parameters $\mu_j, \sigma_{j,t}^2$ are fixed.

One way of obtaining some inferences about the parameters in a NM($K$)-GARCH(1,1) model is based on the asymptotic theory. The standard errors of the parameter estimates are obtained from the inverse of the observed information matrix. However, the only difficulty with this method is the determination of the asymptotic variance of the estimator basing on the large sample theory. Fortunately, in particular for mixture model, it is well known that the sample size $T$ has to be very large before asymptotic theory of maximum likelihood can be applied. Furthermore, Basford et al. (1997) found that unless the sample size was very large, the standard errors found by an information-based approach were too unstable to be recommended for the normal mixture models.

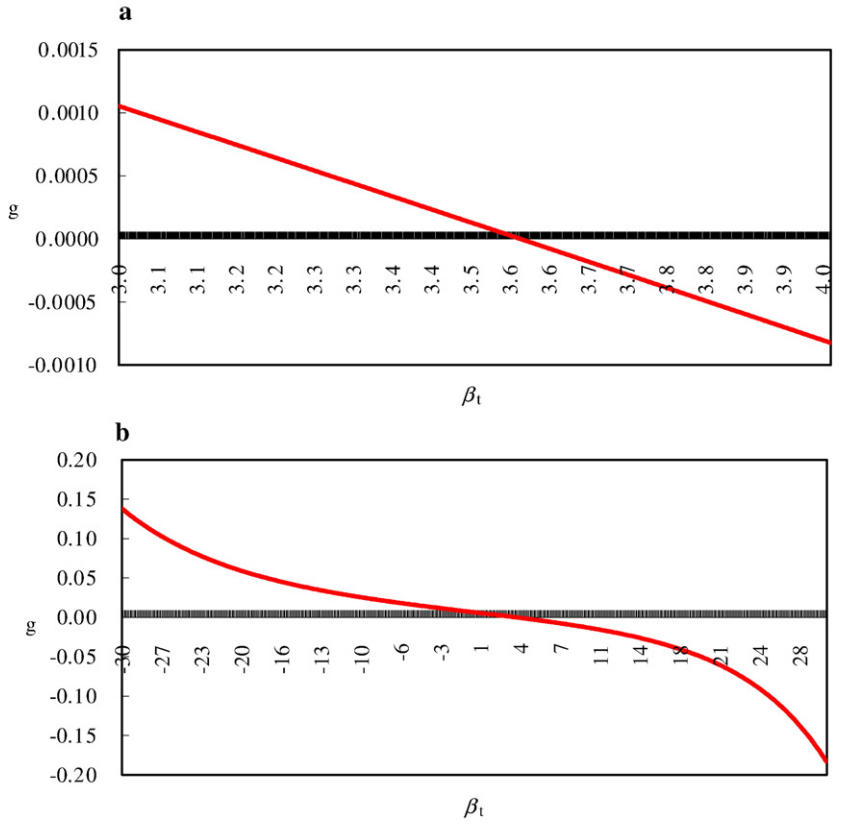


Fig. 1. Graphical determination of the value of RRA, $\beta_t$, for which the equilibrium condition is satisfied, i.e. $g(\beta_t|Z_{t-1}, \mu_j, \sigma^2_{j,t}, j=1,..,K)=0$. The exhibitions are based on $K=2$, $(p_1, p_2)=(0.3, 0.7)$, $(\mu_1, \mu_2)=(-0.014, 0.014)$,and $(\sigma^2_{1,t},\sigma^2_{2,t})=(0.003, 0.001)$.

Therefore, in this paper, we apply the parametric bootstrap method (see, Efron and Tibshirani, 1993) to our empirical studies based on the MLE approach and the algorithm is described as follows.

Step 1 Find the MLE, $\hat{\Psi}^{\mathrm{MLE}}$.
Step 2 Draw $T$ random variables $Y^*_{j1}, Y^*_{j2},..., Y^*_{j,t}$ from the NM(K)-GARCH(1,1) model with the parameter $\hat{\Psi}_{\mathrm{MLE}}$, and use $Y^*_{j1}$, $Y^*_{j2},..., Y^*_{j,t}$ to do inference.
Step 3 Repeat step 2 for $j=1, 2,..., B$.

When $B$ is large enough, we can use the information in step 2 to understand the properties about the estimator based on data, $Y_1$, $Y_2,..., Y_T$. An important index of the precision of a sample-based estimate is the standard deviation (S.D.) of the estimator. This estimated S.D. is the standard deviation estimate computed over the Monte Carlo bootstrap sampling distribution:

$$S.D.(\hat{\beta}) = \sqrt{\sum_{j=1}^{B}\left[\hat{\beta}_j - (1/B)\sum_{j=1}^{B}\hat{\beta}_j\right]^2 \Big/ (B-1)}. \tag{17}$$

Throughout this paper, we replicate $B=50{,}000$ times to make inferences for dynamic RRA estimated by the MLE approach.

### 3.2. Bayesian estimation of RRA

We have seen that estimation for RRA based on NM($K$)-GARCH(1,1) models is straightforward using the EM algorithm. Meanwhile, with the advent of inexpensive and high-speed computers, estimation in Bayesian framework is now feasible using posterior simulation via Markov chain Monte Carlo (MCMC) methods. From past experiences, we would not expect inference about the parameters in Eq. (10) to be highly sensitive to prior specification. In general, we may prefer non-informative priors to informative priors, if no prior information is available. Nevertheless, the main hindrance in normal mixture models is that improper non-informative priors will not yield proper posterior distributions (Diebolt and Robert, 1994; Roeder and Wasserman, 1997). Therefore, in this section, we choose a fixed number of components, $K$, according to BIC and refer to Richardson and Green (1997) to construct weakly informative priors for model parameters. As in Richardson and Green (1997), we assume that $\mu_j$ are drawn independently with normal priors,

$$\mu_j \sim N(\xi, \kappa^{-1}) \tag{18}$$

For the variance processes, we assume priors between the $\theta=(w_j,\ a_j,\ b_j)$ are independent uniform distributions, which are given by

$$\Pi(w,a,b) = \prod_{j=1}^{K} \Pi(w_j,a_j,b_j) \propto \prod_{j=1}^{K} I(w_j>0, a_j\geq 0, b_j\geq 0,\ a_j+b_j<1), \tag{19}$$

where $I(\cdot)$ is the indicator function with $I(S)=1$ if the event $S$ is true, otherwise $I(S)=0$. The prior on the weights $p=(p_1, p_2,..., p_k)$ will always be taken as symmetric Dirichlet,

$$\boldsymbol{p} \sim D(\delta, \delta, \ \ldots, \delta). \tag{20}$$

In order to give weakly informative priors for the model parameters, we introduce hyper-prior and hyper-parameter choices which correspond to making the minimal assumption on the data. Before determining the hyper-parameters, we comment briefly on the issue of labeling the components. The whole model is invariant with respect to permutation of the labels $j=1, 2, ..., K$. For identifiability, Richardson and Green (1997) adopt a unique labeling in which the $\mu_j$ are in increasing numerical order. Hence, the joint prior distribution of the $\mu_j$ is $K!$ times the product of the individual normal densities, restricted to the set $\mu_1<\mu_2<...<\mu_k$. Following Richardson and Green (1997) we take the $N(\xi, \kappa^{-1})$ prior for $\mu_j$ to be rather flat over the range of data, by letting $\xi$ equal to the mid-point of this range, and $\kappa$ equal to a small multiple of $1/R^2$, where $R$ is the length of the range. The complete hierarchical model is displayed in Fig. 2 as a directed acyclic graph (DAG). We follow the usual convention of graphical models that square boxes represent fixed or observed quantities and circles represent the unknowns.

In order to make inferences about model parameters, we need to integrate over high dimensional probability distribution, which could be very difficult. MCMC methods are very

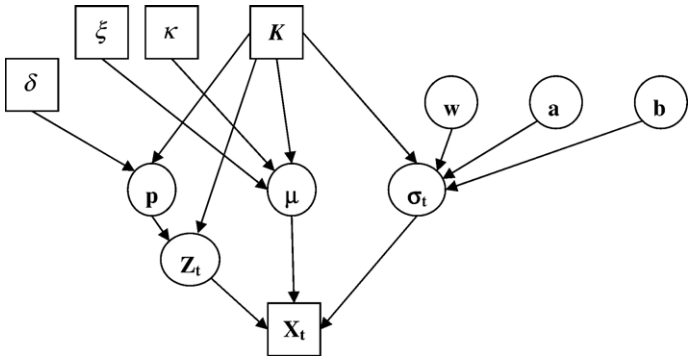


Fig. 2. Directed acyclic graph specific to the complete hierarchical model.

helpful for solving our problem. MCMC is essentially Monte Carlo integration using Markov chains. It draws samples from the required distribution by running a cleverly constructed Markov chain for a long time and then forms sample averages to approximate expectations. The Gibbs sampler and the Metropolis–Hastings (M–H) algorithm are well known among the several ways of constructing those chains. A great advantage of the Gibbs sampler and the M–H algorithm is the ease of implementation which makes heavy use of the modern computational capabilities. Excellent tutorials on the methodology have been provided by Casella and George (1992), Chib and Greenberg (1995), Gilk, Richardson and Spiegelhalter (1996). The MCMC methods are used to make inferences in this section. For the distribution of $\beta_t$ based on hierarchical NM(K)-GARCH(1,1) models, we shall use five move types.

1. Updating the weights $\boldsymbol{p}=(p_1, p_2,\ldots, p_K)$

Through conjugacy, the full conditional for the weights $p$ remains Dirichlet in form:

$$\boldsymbol{p}^{(m)}|\ldots \sim D(\delta + n_1^{(m-1)}, \ldots, \delta + n_K^{(m-1)}), \tag{21}$$

where $n_j^{(m)} = \sum_{t=1}^{T} z_{tj}^{(m)}$ is the number of observations currently allocated to the $j$ component of the normal mixture. Here and the rest of the paper, '|' denotes 'conditional on all other variables'.

2. Updating the parameters $\mu=(\mu_1, \mu_2,\ldots, \mu_K)$

The full conditionals for $\mu_j$ are

$$\mu_j^{(m)}|\ldots \sim N\left(\frac{\sum_{t:z_t^{(m-1)}=j}\left(\frac{x_t}{\sigma_{j,t}^{-2(m-1)}}\right)+\kappa\xi}{\sum_{t:z_t^{(m-1)}=j}\left(\frac{1}{\sigma_{j,t}^{-2(m-1)}}\right)+\kappa}, \left(\sum\nolimits_{t:z_t^{(m-1)}=j}\left(\frac{1}{\sigma_{j,t}^{-2(m-1)}}\right)+\kappa\right)^{-1}\right). \tag{22}$$

In order to preserve the ordering constraint on the $\mu_j$, the move is accepted provided the ordering is unchanged and rejected otherwise.

3. Updating the parameters $(w_j, a_j, b_j)$, $j=1,\ldots, K$.

The posterior conditional density for $(w_j, a_j, b_j)$ are

$$p(w_j^{(m)}, a_j^{(m)}, b_j^{(m)}|\ldots) \propto \prod_{t:z_{tj}^{(m-1)}=1} \phi_j(y_t|Z_{t-1}) \times I(w_j>0, a_j\geq 0, b_j\geq 0, a_j + b_j<1), \tag{23}$$

We update $(w_j, a_j, b_j)$ independently by using the Metropolis–Hastings (MH) algorithm.

4. Updating the allocation $z_{tj}$

For the allocations we have conditional probability

$$f(z_{tj}^{(m)} = 1|\dots) \propto \frac{p_j^{(m)}}{\sigma_j^{(m)}} \exp\left\{-\frac{\left(x_i - \mu_j^{(m)}\right)^2}{2\sigma_j^{2(m)}}\right\}. \tag{24}$$

We can sample directly from this distribution and update the allocation variables independently through Gibbs sampling.

5. Updating RRA $\beta_t$

$\beta_t$ must satisfy the following equation

$$g(\beta_t^{(m)}|\Im_{t-1}, \mu_j^{(m)}, \sigma_{j,t}^{2^{(m)}}, j = 1, \dots, K) = \sum_{j=1}^{K} p_j^{(m)} e^{\mu_j^{(m)}(1-\beta_t^{(m)}) + \frac{\sigma_{j,t}^{2(m)}(1-\beta_t^{(m)})^2}{2}} - \sum_{j=1}^{K} p_j^{(m)} e^{-\mu_j^{(m)}\beta_t^{(m)} + \frac{\sigma_{j,t}^{2(m)}\beta_t^{(m)2}}{2}} = 0. \tag{25}$$

We update $\beta_t$ independently by means of the Gibbs sampler.

The results of Bayesian estimation in this paper correspond to runs of 100,000 sweeps after a burn-in period of 50,000 sweeps. The following settings are used for the previously unspecified constants: $\kappa = 1/R^2$ and $\delta = 1$. Richardson and Green (1997) and Stephens (1997) pointed out that these values convey the belief that "the posterior distributions of parameters are similar, without being informative about their absolute size". Here, equal-tails probability is also used in estimating the 95% posterior interval.

## 4. Simulation studies

We perform simulation studies to compare the proposed MLE and Bayesian RRA estimations based on the NM($K$) model with the GMM method and to determine the numbers of components based on some different information criteria.

In the first part, we generate data from the NM(2)[1] model with five different sample sizes, $N$=50, 100, 200, 500, 1000. The true parameters are $(p_1, p_2)=(0.3,0.7)$, $(\mu_1, \mu_2)=(-0.014, 0.014)$, and $(\sigma_{1,t}^2, \sigma_{2,t}^2)=(0.003, 0.001)$, thus, these values imply the true RRA parameter, $\beta_{\text{True}}=3.564$. Table 1 displays the average and standard deviation of RRA estimators as well as their mean square errors, $\text{MSE}(\hat{\beta})=\text{bias}^2+\text{Var}(\hat{\beta})$, across 10,000 samples. We see that MLE and GMM estimators have moderate upward biases, but Bayesian estimator has mild downward bias for small sample size. However, as the sample size increases the bias is substantially reduced for all estimation methods. We also find that the RRA estimator based on the NM(2)-GARCH(1,1) model are more efficient than that based on the GMM model. In addition, based on the MSE criterion, the RRA estimated by MLE is similar to that estimated by Bayesian approach, but they are slightly more accurate than GMM estimators in large samples. In small samples, RRA estimated by the Bayesian approach with a weakly informative prior is obviously more accurate than GMM and MLE.

In the second part, the AIC and BIC criteria are applied to two time series generated by the NM (2)-GARCH(1,1) model with different sample sizes, $N$=200 and 500. We set the true parameter

[1] If we generate data from the NM(2)-GARCH(1,1) model, we will only have the true parameter values of GARCH process but the true values of the volatilities at time $t$. This will produce forecasted RRA but true RRA, and the comparisons may be meaningless. This is why we do not simulate data from the NM(2)-GARCH(1,1) model but from the NM(2) model.

Table 1
Simulation results from MLE, Bayesian, and GMM

| $N$ | $\beta_{\text{True}}$ | MLE | | | Bayesian | | | GMM | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\hat{\beta}$ | SE($\hat{\beta}$) | MSE($\hat{\beta}$) | $\hat{\beta}$ | SE($\hat{\beta}$) | MSE($\hat{\beta}$) | $\hat{\beta}$ | SE($\hat{\beta}$) | MSE($\hat{\beta}$) |
| 50 | 3.564 | 4.069 | 3.907 | 15.520 | 3.390 | 3.798 | **14.455** | 4.036 | 4.043 | 16.569 |
| 100 | 3.564 | 3.808 | 2.626 | 6.955 | 3.433 | 2.508 | **6.307** | 3.809 | 2.633 | 6.993 |
| 200 | 3.564 | 3.637 | 1.797 | 3.235 | 3.482 | 1.707 | **2.921** | 3.682 | 1.799 | 3.250 |
| 500 | 3.564 | 3.626 | 1.094 | **1.201** | 3.484 | 1.095 | 1.205 | 3.626 | 1.099 | 1.212 |
| 1000 | 3.564 | 3.586 | 0.771 | **0.595** | 3.548 | 0.772 | 0.596 | 3.592 | 0.779 | 0.608 |

The table considers $N=50$, 100, 200, 500, and 1000 observations obtained by generating sample from NM(2) models with true parameter values ( $p_1$, $p_2$)=(0.3, 0.7), ($\mu_1$, $\mu_2$)=(−0.014, 0.014),and ($\sigma^2_{1,t}$, $\sigma^2_{2,t}$)=(0.003, 0.001) . Therefore, from Eq. (15), we have the true RRA, $\beta_{\text{True}}=3.564$ By simulating 10,000 times, the means, $\hat{\beta}$, standard deviations, SE($\hat{\beta}$), and mean square errors, MSE($\hat{\beta}$), of RRA estimators for three different estimating methods are shown. The bold figure represents the minimum MSE among MLE, Bayesian, and GMM methods.

values to be ($p_1$, $p_2$)=(0.3, 0.7), ($\mu_1$, $\mu_2$)=(−0.014, 0.014), ($w_1$,$a_1$,$b_1$)=(0.00015, 0.1, 0.85) , and ($w_1$,$a_1$,$b_1$)=(0.00005, 0.1, 0.85). Then, we use the EM algorithm to find the MLE estimators of those parameters and calculate the maximum log conditional likelihood, $\ell$, AIC and BIC values. Table 2 shows that the AIC criterion tends to select too many components and the BIC criterion correctly selects the true number of components.

## 5. Empirical studies

The empirical analyses in this paper are based on monthly market rates of return and riskless rates during the period January 1941 through December 2001 with a sample of $T=732$ observations. The value-weighted index of the New York Stock Exchange from 1941 to 2001 are collected from the Center for Research in Security Prices (CRSP) at the University of Chicago and used as the proxies for the rates of return on the market portfolio. The proxies for the risk-free rates are monthly returns of U.S. Treasury bills in stock, bonds, and inflation. Using this set of data, we calculate the excess return $x=(1+R_{mt})/(1+R_{ft})$, the return in the market portfolio over the risk-free rate, for doing the following empirical studies.

Table 3 provides the summary statistics for the whole period of the log excess returns which are based on the monthly value-weighted index and the 30-day U.S. Treasury bills returns from January

Table 2
Simulation likelihood-based goodness of fit

| $K$ | $d$ | $N=200$ | | | $N=500$ | | |
|---|---|---|---|---|---|---|---|
| | | $\ell$ | AIC | BIC | $\ell$ | AIC | BIC |
| 1 | 4 | 358.30 | −708.60 | −695.41 | 924.12 | −1840.24 | −1827.05 |
| 2 | 9 | 387.78 | −757.55 | **−727.87** | 1003.34 | −1988.68 | **−1959.00** |
| 3 | 15 | 400.84 | **−771.68** | −722.20 | 1009.85 | **−1989.70** | −1940.23 |
| 4 | 20 | 405.25 | −770.49 | −704.52 | 1010.97 | −1981.94 | −1915.97 |

The table considers $N=200$ and $N=500$ observations obtained by generating sample from NM(2)-GARCH(1,1) with true parameter values ( $p_1$, $p_2$)=(0.3, 0.7), ($\mu_1$, $\mu_2$)=(−0.014, 0.014), ($\omega_1$, $\alpha_1$, $\beta_1$)=(0.00015,0.1, 0.85) and ($\omega_2$, $\alpha_2$, $\beta_2$)=(0.00005,0.1, 0.85). The columns labeled $K$ and $d$ refer to the number of required component densities and the number of parameters for the respective model; $\ell$ is the log likelihood; AIC=$-2\ell+2d$; BIC=$-2\ell+d\log N$. For each of the three criteria, the criterion values are shown for $K=1$, 2, 3, 4. The bold figure represents the optimal model, which is chosen by AIC or BIC criteria.

Table 3
Descriptive statistics for monthly log excess returns (%) of value-weighted index (1/1941–12/2001)

| A. Summary statistics for log excess returns (%) | | | |
|---|---|---|---|
| Mean | 0.5957 | S.D. | 4.2332 |
| Minimum | −25.8077 | Skewness | −0.7253 |
| Median | 0.9897 | Kurtosis | 5.5227 |
| Maximum | 14.7679 | | |
| **B. Test for Normality** | | | |
| Jarque–Bera test | | Chi-square test | |
| J–B stat. | 258.2843*** | Chi-square stat. | 51.3689*** |
| J–B *p*-value | <0.001 | Chi-square *p*-value | <0.001 |
| **C. Ljung–Box test for log excess returns, $y_t$** | | | |
| $Q(5)$ stat. | 9.0876 | $Q(5)$ *p*-value | 0.106 |
| $Q(10)$ stat. | 11.2520 | $Q(10)$ *p*-value | 0.338 |
| **D. Ljung –Box test for square of log excess returns, $y_t^2$** | | | |
| $Q(5)$ stat. | 9.5489* | $Q(5)$ *p*-value | 0.089 |
| $Q(10)$ stat. | 17.2170* | $Q(10)$ *p*-value | 0.070 |

The table shows the summary statistics of log(monthly excess returns), $y_t$, over the period from 1/1941 to 12/2001 in panel A. The statistics and *p*-value of chi-square and Jarque and Bera (1980) tests are used to check the closeness of the data to a normal density in panel B. Moreover, we exhibit the *p*-value of Ljung-Box *Q*-statistic (1978) to measure the serial correlations of *yt* and $yt^2$ using five and ten lagged values. *, ** and *** indicate statistical significance at the 10%, 5% and 1% levels (two tailed test), respectively.

1941 to December 2001. The summary statistics includes sample mean, minimum, median, maximum, standard deviation, skewness, kurtosis, the *p*-values of chi-square and Jarque and Bera (1980)[2] tests to check the normality of the data, and the *p*-values of Ljung and Box (1978) to verify the serial correlations of returns and square of returns data. From Table 3, we can observe that the unconditional distribution of the log excess return has negative skewness and heavy tails relative to the normal distribution. For chi-square and Jarque and Bera (1980) tests, we reject the normality assumption for overall periods at 1% significant level. Therefore, it appears that normal distribution is not an adequate assumption to the log transformation for the excess returns. In addition, the results of Ljung-Box test confirm that log monthly excess returns of the value-weighted index for overall period have no significant serial correlations. Table 3 also shows that there are some serial correlations for square of the log returns and there is volatility clustering phenomenon. Therefore, assuming log monthly excess returns to be NM(*K*)-GARCH(1,1) model may be more appropriate.

We set $K=2$ according to the BIC criterion. Therefore, the MLE and Bayesian estimators and 95% confidence intervals of dynamic RRA based on the NM(2)-GARCH(1,1) model from 1/1956 to 12/2001 are shown in Figs. 3 and 4[3]. We find that the coefficients of RRA are not stable

[2] The Jarque–Bera test for normality is based on skewness (Skew) and kurtosis (Kurt). The statistic is given by $\mathrm{JB} = \frac{T}{6}\left(Skew^2 + \frac{(\mathrm{Kurt}-3)^2}{4}\right)$, which has a chi-squared distribution with 2 degrees of freedom.

[3] It is well known that the structure of the market may vary a lot across different periods. If we estimate the RRA based on the whole sample period, we will only mildly capture the variation in the market structure. Therefore, in our analyses, we estimate the RRA at time *t* based on the data from time $t-180$ to $t-1$. Although, we can also apply the BIC criterion to choose different number of components, *K*, across the rolling sample. However, we think that this operation may be worthless. Therefore, we determine *K* only by basing on the whole sample and adopt this result to do the following analyses.

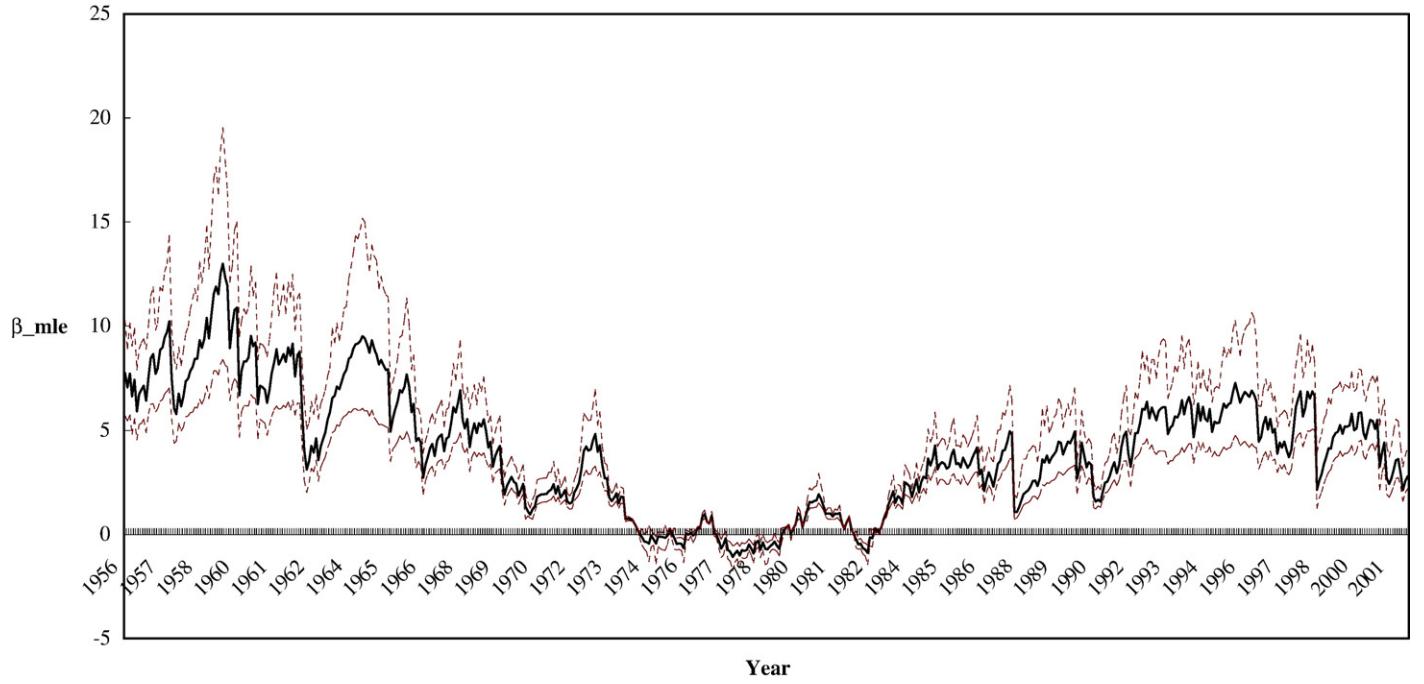


Fig. 3. The MLE estimators of dynamic RRA (solid line) and 95% confidence intervals (dotted line) in the NM(2)-GARCH(1,1) model, when proxy for the rates of return on the market portfolio are the CRSP value weighted index during 1/1956 through 12/2001.

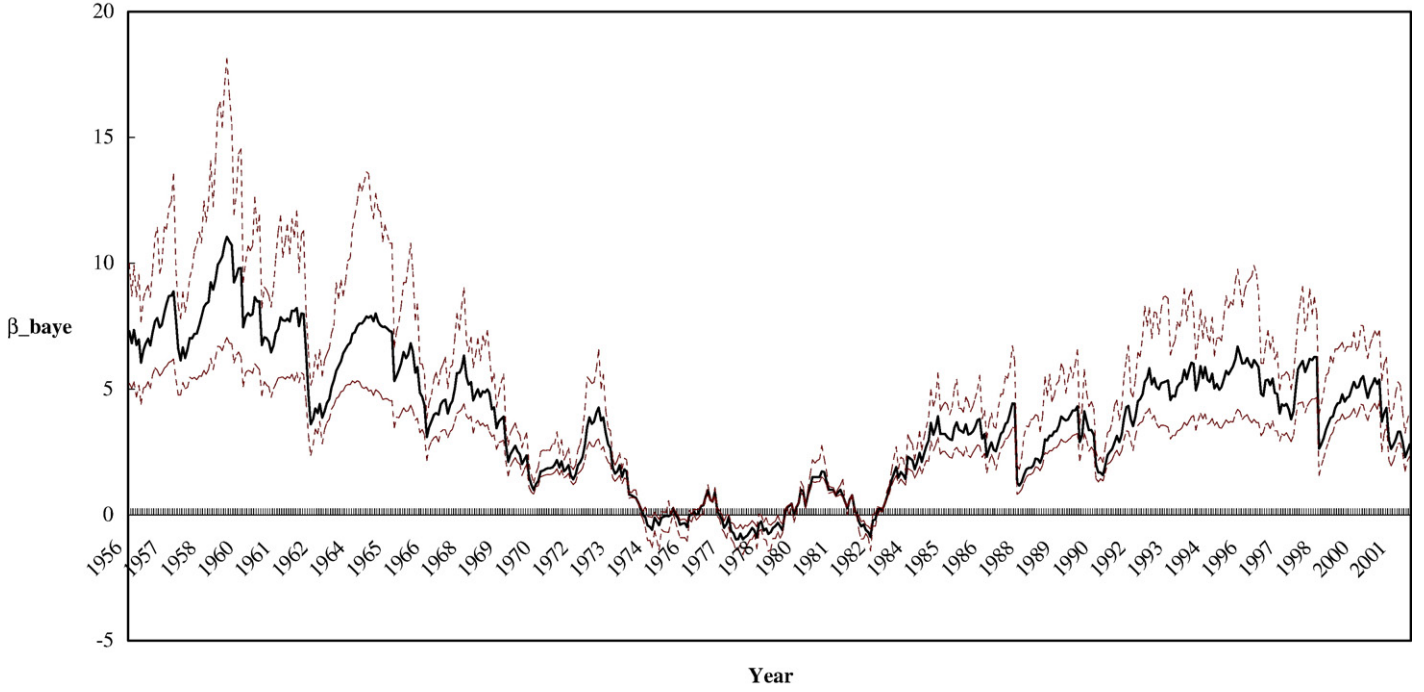


Fig. 4. The Bayesian estimators of dynamic RRA (solid line) and 95% posterior intervals (dotted line) in the NM(2)-GARCH(1,1) model, when proxy for the rates of return on the market portfolio are the CRSP value weighted index during 1/1956 through 12/2001.

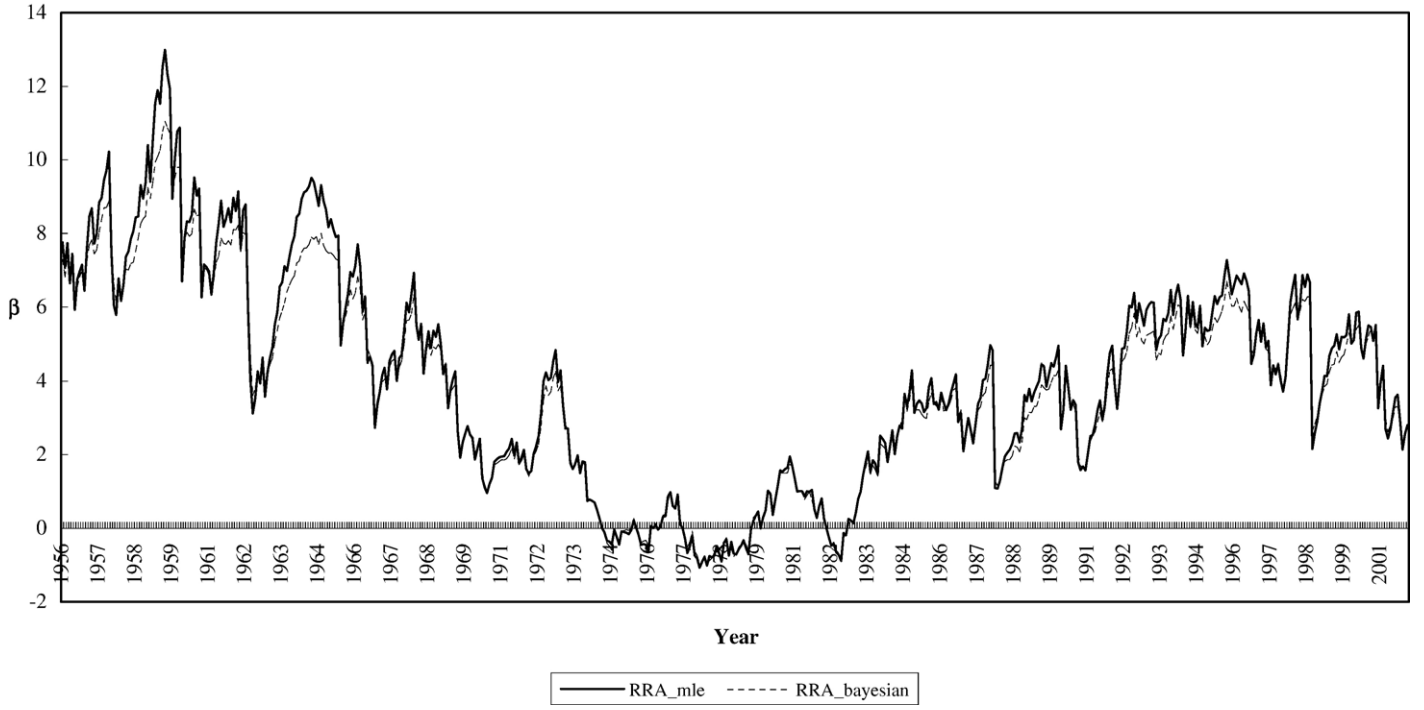


Fig. 5. The comparisons between MLE and Bayesian estimators of dynamic RRA in the NM(2)-GARCH(1,1) model, when proxy for the rates of return on the market portfolio are the CRSP value weighted index during 1/1956 through 12/2001.

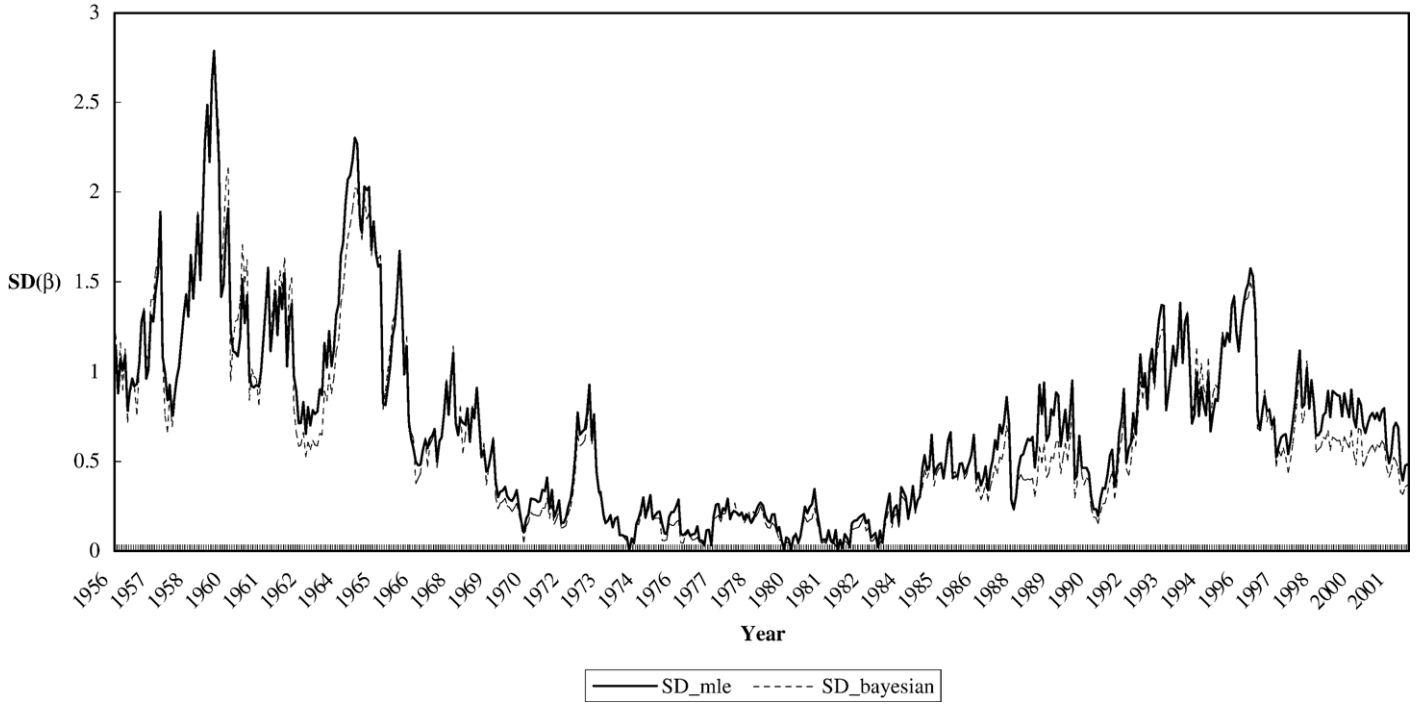


Fig. 6. The comparisons between standard deviations of MLE and Bayesian estimators of dynamic RRA in the NM(2)-GARCH(1,1) model, when proxy for the rates of return on the market portfolio are the CRSP value weighted index during 1/1956 through 12/2001.

throughout overall period. Largest and smallest RRA are in the 1950's and 1970's, respectively. The coefficients of RRA tend to decline from 1950's to 1970's, and increase from 1980's to 1990's. Figs. 5 and 6 exhibit the RRA estimators and their standard error comparisons between MLE and Bayesian methods. The RRA estimated by MLE method, on average, is larger than that estimated by Bayesian method. In addition, we find that Bayesian estimation method is more efficient than MLE estimation method. In Fig. 6, we also see that the volatilities of RRA are larger before 1974 and after 1983 but lower between 1974 and 1983.

## 6. Conclusions

This paper shows that the utility-based model of asset pricing can be estimated to be dynamic and more accurate with the NM($K$)-GARCH(1,1) model for log asset returns. Then, the MLE method with the EM algorithm and the Bayesian approach with a weakly informative prior are derived to estimate RRA. The empirical findings are as follows. First, the log excess returns can be adequately characterized by the NM($K$)-GARCH(1,1) model with their conditional and unconditional nonzero sknewness and excess kurtosis. Secondly, we have identified that RRA estimator is statistically efficient with the robust model assumption. Third, the RRA estimator obtained by the Bayesian approach with a weakly informative prior performs better in small samples based on the MSE criterion. Finally, Bayesian approach can combine an investor's prior belief about the accuracy of the pricing model and the information in the data and describe the sampling distribution of RRA estimator.

## Appendix A

Proof that $\beta_t$ *is a continuous function of* $(p_j, \mu_j, \sigma^2_{j,t})$ *for* $j=1, 2, \ldots, K$.
Consider

$$g(\beta_t|\Im_{t-1}, \mu_j, \sigma^2_{j,t}, j=1, \ldots, K) = \sum_{j=1}^{K} p_j e^{\mu_j(1-\beta_t)+\frac{\sigma^2_{j,t}(1-\beta_t)^2}{2}} - \sum_{j=1}^{K} p_j e^{-\mu_j\beta_t+\frac{\sigma^2_{j,t}\beta_t^2}{2}} = 0$$

which has a unique solution when the other parameters $p_j$, $\mu_j$, $\sigma^2_{j,t}$=1, 2,..., $K$ are given. We have the following first derivative of $g$,

$$g'(\beta_t|\Im_{t-1}, \mu_j, \sigma^2_{j,t}, j=1, \ldots, K) = \sum_{j=1}^{K} p_j \left[(-\mu_j + \sigma^2_{j,t}(\beta_j - 1)) \cdot e^{\mu_j(1-\beta_t)+\frac{\sigma^2_{j,t}(1-\beta_t)^2}{2}} - (-\mu_j + \sigma^2_{j,t}\beta_t) \cdot e^{-\mu_j\beta_t+\frac{\sigma^2_{j,t}\beta_t^2}{2}}\right].$$

It is easy to see that

$$e^{\mu_j(1-\beta_t)+\frac{\sigma^2_{j,t}(1-\beta_t)^2}{2}} - e^{-\mu_j\beta_t+\frac{\sigma^2_{j,t}\beta_t^2}{2}} \begin{cases} <0 & \text{if} \quad \beta_t > \dfrac{\mu_j}{\sigma^2_{j,t}} + \dfrac{1}{2} \\ = 0 & \text{if} \quad \beta_t = \dfrac{\mu_j}{\sigma^2_{j,t}} + \dfrac{1}{2} \\ >0 & \text{if} \quad \beta_t < \dfrac{\mu_j}{\sigma^2_{j,t}} + \dfrac{1}{2} \end{cases} \text{for } j = 1, 2, \ldots, K.$$

For each $j$, when $\beta_t > \frac{\mu_j}{\sigma_{j,t}^2} + \frac{1}{2}$, we have

$$-\mu_j + \sigma_{j,t}^2\beta_t > \mu_j + \sigma_{j,t}^2\left(\frac{\mu_j}{\sigma_{j,t}^2} + \frac{1}{2}\right)$$

$$= \frac{1}{2}\sigma_{j,t}^2 > 0, \text{ and thus } \left[(-\mu_j + \sigma_{j,t}^2(\beta_j - 1)) \cdot e^{\mu_j(1-\beta_t) + \frac{\sigma_{j,t}^2(1-\beta_t)^2}{2}} - (-\mu_j + \sigma_{j,t}^2\beta_t) \cdot e^{-\mu_j\beta_t + \frac{\sigma_{j,t}^2\beta_t^2}{2}}\right]$$

$$\leq \left[(-\mu_j + \sigma_{j,t}^2\beta_j) \cdot e^{\mu_j(1-\beta_t) + \frac{\sigma_{j,t}^2(1-\beta_t)^2}{2}} - (-\mu_j + \sigma_{j,t}^2\beta_t) \cdot e^{-\mu_j\beta_t + \frac{\sigma_{j,t}^2\beta_t^2}{2}}\right]$$

$$= \left[(-\mu_j + \sigma_{j,t}^2\beta_j) \cdot \left(e^{\mu_j(1-\beta_t) + \frac{\sigma_{j,t}^2(1-\beta_t)^2}{2}} - e^{-\mu_j\beta_t + \frac{\sigma_{j,t}^2\beta_t^2}{2}}\right)\right] < 0.$$

Alternatively, when $\beta_t < \frac{\mu_j}{\sigma_{j,t}^2} + \frac{1}{2}$,

$$-\mu_j + \sigma_{j,t}^2\beta_t - \sigma_{j,t}^2 < -\mu_j + \sigma_{j,t}^2\left(\frac{\mu_j}{\sigma_{j,t}^2} + \frac{1}{2}\right) - \sigma_{j,t}^2 = -\frac{1}{2}\sigma_{j,t}^2 < 0.$$

So, we have

$$\left[(-\mu_j + \sigma_{j,t}^2(\beta_j - 1)) \cdot e^{\mu_j(1-\beta_t) + \frac{\sigma_{j,t}^2(1-\beta_t)^2}{2}} - (-\mu_j + \sigma_{j,t}^2\beta_t) \cdot e^{-\mu_j\beta_t + \frac{\sigma_{j,t}^2\beta_t^2}{2}}\right]$$

$$\leq \left[(-\mu_j + \sigma_{j,t}^2\beta_t - \sigma_{j,t}^2) \cdot e^{\mu_j(1-\beta_t) + \frac{\sigma_{j,t}^2(1-\beta_t)^2}{2}} - (-\mu_j + \sigma_{j,t}^2\beta_t - \sigma_{j,t}^2) \cdot e^{-\mu_j\beta_t + \frac{\sigma_{j,t}^2\beta_t^2}{2}}\right]$$

$$= \left[(-\mu_j + \sigma_{j,t}^2\beta_j - \sigma_{j,t}^2) \cdot \left(e^{\mu_j(1-\beta_t) + \frac{\sigma_{j,t}^2(1-\beta_t)^2}{2}} - e^{-\mu_j\beta_t + \frac{\sigma_{j,t}^2\beta_t^2}{2}}\right)\right] < 0.$$

Hence, for $\beta_t \varepsilon \Re$ and $j = 1, 2, \ldots, K$,

$$\left[(-\mu_j + \sigma_{j,t}^2(\beta_t - 1)) \cdot e^{\mu_j(1-\beta_t) + \frac{\sigma_{j,t}^2(1-\beta_t)^2}{2}} - (-\mu_j + \sigma_{j,t}^2\beta_t) \cdot e^{-\mu_j\beta_t + \frac{\sigma_{j,t}^2\beta_t^2}{2}}\right] \leq 0.$$

Accordingly, as long as $\frac{\mu_j}{\sigma_{j,t}^2} + \frac{1}{2}$, are not identical for $j = 1, 2, \ldots, K$, then $g'(\beta_t)$ is strictly negative. Furthermore, because

$$\lim_{\beta_t \to \infty}\left(\sum_{j=1}^{K} p_j\left(e^{\mu_j(1-\beta_t) + \frac{\sigma_{j,t}^2(1-\beta_t)^2}{2}} - e^{-\mu_j\beta_t + \frac{\sigma_{j,t}^2\beta_t^2}{2}}\right)\right) = \infty,$$

$$\lim_{\beta_t \to \infty}\left(\sum_{j=1}^{K} p_j\left(e^{\mu_j(1-\beta_t) + \frac{\sigma_{j,t}^2(1-\beta_t)^2}{2}} - e^{-\mu_j\beta_t + \frac{\sigma_{j,t}^2\beta_t^2}{2}}\right)\right) = -\infty,$$

and $g(\beta_t)$ is continuous, $g(\beta_t) = 0$ has at least one real solution.

For these reasons, $g(\beta_t)$ is a strictly decreasing continuous function and $g(\beta_t) = 0$ has a unique solution. Therefore, we can consider $\beta_t$ a continuous function of $p_j, \mu_j, \sigma_j^2$ for $j = 1, 2, \ldots, K$.

## References

Aitkin, M., Tunnicliffe Wilson, G., 1980. Mixture models, outliers, and the EM algorithm. Technometrics 22, 325–331.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 267–281.

Alexander, C., Lazar, E., 2005. Normal Mixture GARCH(1,1): application to exchange rate modelling. Journal of Applied Econometrics 20, 1–30.
Alonso, A., Rubio, G., Tusell, F., 1990. Asset pricing and risk aversion in the Spanish stock market. Journal of Banking and Finance 14, 351–369.
Bansal, R., Viswanathan, S., 1993. No arbitrage and arbitrage pricing — a new approach. Journal of Finance 48, 1231–1262.
Basford, K.E., Greenway, D.R., McLachlan, G.J., Peel, D., 1997. Standard errors of fitted means under normal mixture models. Computational Statistics 12, 1–17.
Boothe, P., Glassman, D., 1987. The statistical distribution of exchange rates: empirical evidence and economic implications. Journal of International Economic 22, 297–319.
Brown, D.P., Gibbons, M.R., 1985. A simple econometric approach for utility-based asset pricing model. Journal of Finance 40, 359–381.
Campbell, J.Y., 1993. Intertemporal asset pricing without consumption data. American Economic Review 83, 487–512.
Casella, G., George, E.I., 1992. Explaining the gibbs sampler. American Statistician 46, 167–174.
Chib, S., Greenberg, E., 1995. Understanding the Metropolis–Hastings algorithm. American Statistics 49, 327–335.
Cox, J., Ingersoll, J., Ross, S., 1985. The theory of the term structure of interest rates. Econometrica 53, 385–407.
Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B 39, 1–38.
Diebolt, J., Robert, C.P., 1994. Estimation of finite mixture distributions through Bayesian sampling. Journal of the Royal Statistical Society. Series B 56, 363–375.
Duan, K.B., Singleton, K.J., 1986. Modelling the term structure of interest rates under habit formation and durability of goods. Journal of Financial Economics 17, 27–55.
Efron, B., Tibshirani, R., 1993. An Introduction to the Bootstrap. Chapman and Hall, London.
Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987–1006.
Ermini, L., 1989. Some new evidence on the timing of consumption decisions and their generating process. Review of Economics and Statistics 71, 643–650.
Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. Markov chain Monte Carlo in Practice. Chapman and Hall, London.
Grossman, S., Shiller, R., 1981. The determinants of variability of stock market prices. American Economic Review 71, 222–227.
Haas, M., Mittnik, S., Paolella, M., 2004. Mixed normal conditional heteroskedasticity. Journal of Financial Econometrics 2, 211–250.
Hakansson, N., 1970. Optimal investment and consumption under risk for a class of utility functions. Econometrica 38, 587–607.
Hansen, L.P., Jagannathan, R., 1991. Implications of security market data for models of dynamic economies. Journal of Political Economy 99, 225–262.
Hansen, L.P., Singleton, K.J., 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica 50, 1269–1286.
Hansen, L.P., Singleton, K.J., 1983. Stochastic consumption, risk aversion and the temporal behavior of asset returns. Journal of Political Economy 91, 249–265.
Jarque, C.M., Bera, A.K., 1980. Efficient tests for normality, homoscedasticity, and serial independence of regression residuals. Economics Letters 6, 255–259.
Karson, M., Cheng, D., Lee, C.F., 1995. Sampling distribution of the relative risk aversion estimator: theory and applications. Review of Quantitative Finance and Accounting 5, 43–54.
Kon, S.J., 1984. Models of stock returns: a comparison. Journal of Finance 39, 147–165.
Kraus, A., Litzenberger, R., 1975. Market equilibrium in a multiperiod state preference model with logarithmic utility. Journal of Finance 30, 1213–1227.
Lee, C.F., Lee, J.C., Ni, H.F., Wu, C.C., 2004. On a simple econometric approach for utility-based asset pricing model. Review of Quantitative Finance and Accounting 22, 331–344.
Ljung, G., Box, G.E.P., 1978. On a measure of lack of fit in time series models. Biometrika 66, 67–72.
Lucas, R., 1978. Asset prices in an exchange economy. Econometrica 46, 1429–1445.
McLachlan, G.J., Peel, D., 2000. Finite Mixture Models. John Wiley and Sons, New York.
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.
Pastor, L., Stambaugh, R.F., 2000. Comparing asset pricing models: an investment perspective. Journal of Financial Economics 56, 335–381.

Richardson, S., Green, P.J., 1997. On Bayesian analysis of mixture with an unknown number of components. Royal Statistical Society 59, 731–792.
Roeder, K., Wasserman, L., 1997. Practical density estimation using mixtures of normals. Journal of the American Statistical Association 92, 894–902.
Rosenberg, J.V., Engle, R.F., 2002. Empirical pricing kernels. Journal of Financial Economics 64, 341–372.
Rothschild, M., Stiglitz, J., 1971. Increasing risk II: its economic consequences. Journal of Economic Theory 3, 66–84.
Rubinstein, M., 1976. The valuation of uncertain income streams and the pricing of options. Bell Journal of Economics 7, 407–425.
Rubinstein, M., 1977. The strong case for the generalized logarithmic utility models as the premier model of financial markets. In: Levy, H., Sarnat, M. (Eds.), Financial Decision Making under Uncertainty. Academic Press, New York.
Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.
Slesnick, D.T., 1998. Are our data relevant to the theory? The case of aggregate consumption. Journal of the American Statistical Association 16, 52–61.
Stephens, M., 1997. Discussion of "On Bayesian analysis of mixtures with an unknown number of components". In: Richardson, S., Green, P.J. (Eds.), J. Roy. Statist. Soc. Ser. B, vol. 59, pp. 768–769.
Titterington, D.M., Smith, A.F.M., Makov, U.E., 1985. Statistical Analysis of Finite Mixture Distributions. Wiley, New York.
Wilcox, D., 1992. The construction of U.S. consumption data: some facts and their implications for empirical work. American Economic Review 82, 922–941.
Zehna, P., 1966. Invariance of maximum likelihood estimation. Annals of Mathematical Statistics 37, 744.