

CIRJE-J-173

## MCMC法とその確率的ボラティリティ変動モデルへの応用

東京大学大学院経済学研究科  
大森裕浩

一橋大学経済研究所  
渡部敏明

2007年3月

CIRJE ディスカッションペーパーの多くは  
以下のサイトから無料で入手可能です。  
[http://www.e.u-tokyo.ac.jp/cirje/research/03research02dp\\_j.html](http://www.e.u-tokyo.ac.jp/cirje/research/03research02dp_j.html)

このディスカッション・ペーパーは、内部での討論に資するための未定稿の段階にある論文草稿である。著者の承諾なしに引用・複写することは差し控えられる。

## **Abstract**

In the time series analysis of asset prices, the stochastic volatility models have recently attracted attentions of many researchers since it clearly describes time-varying variance of asset returns. However, it is difficult to evaluate the likelihood and obtain the maximum likelihood estimators of parameters for such models. We take Bayesian approach and use Markov chain Monte Carlo (MCMC) method to overcome such a problem. We first describe MCMC method and conduct a survey of the literature for its application to the stochastic volatility model. The empirical analysis of stock returns data is also given.

# MCMC法とその確率的ボラティリティ変動モデルへの応用

大森裕浩（東京大学大学院経済学研究科）

渡部敏明（一橋大学経済研究所）

## 概要

資産価格の時系列分析では、近年、ボラティリティと呼ばれる価格変化率の2次のモーメントの変動に注目が集まっている。ボラティリティの変動を明示的に定式化する時系列モデルの1つに確率的ボラティリティ変動 (Stochastic Volatility; SV) モデルがある。このモデルは尤度を評価するのが難しいため、パラメータの推定に最尤法に代る推定法が必要になる。そうした推定法の中で注目を集めているものにマルコフ連鎖モンテカルロ (Markov-chain Monte Carlo; MCMC) 法を用いたベイズ推定法があり、この方法は近年めざましい発展を遂げている。本稿では、そうしたMCMC法とそのSVモデルへの応用について解説を行う。推定法の開発に伴って、SVモデル自体も改良が行われるようになってきている。そこで、本稿では、SVモデルの最近の発展についてもサーベイを行うとともに、そうした拡張SVモデルの実際の株式収益率データに応用した結果も紹介する。

## 1 はじめに

近年、資産価格の時系列分析では、ボラティリティと呼ばれる2次のモーメントの変動に注目が集まっている。ボラティリティは投資リスクの指標であるとともに、オプション価格の決定要因でもある。そこで、もしそれが時間を通じて変動するのであれば、その変動をうまく捉えられるような時系列モデルを開発することは、単に学術的な観点からだけでなく、投資のリスク管理という観点から実務家にとっても重要である。ボラティリティの変動を明示的に定式化する時系列モデルとしては、これまでに、大きく分けて、2つのものが提案されている。1つは、Engle (1982) によって提案された ARCH (Autoregressive Conditional Heteroskedasticity) モデルとそれを発展させたモデル (本論文では、以下、そうしたモデルを総称して、ARCH 型モデルと呼ぶ) であり、<sup>1</sup> もう1つは、確率的ボラティリティ変動 (Stochastic Volatility; 以下、略して、SV) モデルである。ARCH 型モデルがパラメータの値を最尤法によって簡単に推定できるのに対して、SV モデルは尤度を解析的に評価するのが難しいため、パラメータの推定には最尤法に代わる推定法が必要になる。

SV モデルの推定法にはこれまでさまざまな方法が提案されているが、その中で、特に注目を集めているものに、Jacquier, Polson and Rossi (1994) によって提案されたマルコフ連鎖モンテカルロ (Markov-chain Monte Carlo; 以下、略して、MCMC) 法を用いたベイズ推定法がある。<sup>2</sup> SV モデルのような尤度を解析的に求められないモデルの場合、パラメータの事後分布もベイズの定理を用いて解析的に求めることができない。そうしたモデルをベイズ推定する場合には、何らかの方法を用いてパラメータの値を事後分布からサンプリングし、サンプリングされた値を用いてパラメータを推定するという方法がとられる。解析的に求められない未知の事後分布からのサンプリングを可能にしてくれるのが、MCMC 法である。<sup>3</sup> MCMC 法とは1回前にサンプリングされた値に基づいて次の値をサンプリングする方法の総称であり、代表的なものにギブス・サンプラーと Metropolis-Hastings (MH) アルゴリズムがある。本稿ではまずこうした MCMC 法について解説を行う。

SV モデルをベイズ推定する場合、パラメータだけでなく潜在変数であるボラティリティも同時事後分布からサンプリングする。その際、ボラティリティは標本の大きさだけあるので、それをいかに効率的にサンプリングするかがポイントとなる。Jacquier, Polson and Rossi (1994) は、各期各期のボラティリティを別々にサンプリングする single-move sampler と呼ばれる方法を用いていたが、この方法を用いるとサンプリングされた値に高い自己相関が生じ、MCMC 法の収束が遅い上、推定値の標準誤差が大きくなるのが Shephard and Pitt (1997) によって示されている。そこで、その後、Shephard and Pitt (1997), Watanabe and Omori (2004a,b) らによって multi-move sampler, Kim, Shephard, and Chib (1998) によって mixture sampler と呼ばれるより効率的なボラティリティのサンプリング法が提案されている。本稿では、こうした MCMC 法を用いた SV モデルのベイズ推定法の最近の発展についてサーベイを行っている。

SV モデルは推定が難しいため、ARCH 型と比べると、これまでモデルの拡張はあまりなされてこなかった。しかし、近年、推定法の開発に伴い、SV モデルの拡張も行われるようになってき

<sup>1</sup>ARCH 型モデルについて詳しくは、Bollerslev, Engle and Nelson (1994), 渡部 (2000) 等を参照のこと。

<sup>2</sup>その他の推定法については、Ghysels, Harvey and Renault (1996), Shephard (2004), 渡部 (2000) を参照のこと。

<sup>3</sup>MCMC 法について詳しくは、大森 (2001), 中妻 (2003), 和合 (2005) を参照のこと。

た. そこで, 本稿では, そうした SV モデル自体の最近の発展についてもサーベイを行っている.

本稿の以下の構成は次の通りである. まず, 第 2 節で MCMC 法について解説を行う. 次に, 第 3 節で通常用いられる最も簡単な SV モデルを紹介し, 続く第 4 節で, そうした簡単な SV モデルの MCMC 法を用いたベイズ推定法を解説する. さらに第 5 節で, SV モデルの最近の発展についてサーベイし, 第 6 節で実際の株式収益率データを用いた推定結果を紹介する. 最後に第 7 節で, 今後の課題について述べる.

## 2 マルコフ連鎖モンテカルロ法

マルコフ連鎖モンテカルロ法は乱数発生を行うモンテカルロ法の 1 つでマルコフ連鎖を利用する点が特徴的である. 基礎的なモンテカルロ法では独立な確率標本を発生させるが, 多変量分布からの確率変数のサンプリングは困難であることも多い. このような場合においてもマルコフ連鎖モンテカルロ法は, サンプリングの効率性という点ではやや劣るものの, 適当なマルコフ連鎖を反復することによって容易に目標分布からの確率標本を発生することができる.

この方法は, マルコフ連鎖の「適当な初期値から始めて十分な回数の連鎖の反復をしていくとき, その確率標本の分布がある条件の下で不変分布に収束する」という性質を利用する. つまり, 不変分布が目標分布になるようにマルコフ連鎖を構成することにより, マルコフ連鎖の確率標本を目標分布からの確率標本とするのである.

マルコフ連鎖モンテカルロ法のアルゴリズムは, メトロポリス - ヘイスティングス (Metropolis-Hastings) アルゴリズムである. このアルゴリズムは, まず Metropolis *et al.* (1953) によって統計物理学の分野で提案されたメトロポリス・アルゴリズムが, Hastings (1970) によって一般化されたものであり, 特別な場合としてよく知られているアルゴリズムにギブス・サンプラー (Gibbs sampler) がある.

### 2.1 マルコフ連鎖

不変分布  $\pi$  をもつ斉時的なマルコフ連鎖とはその推移核  $P$  が, すべての可測集合  $A$  について

$$P(X_n, A) = P\{X_{n+1} \in A | X_0, \dots, X_n\}, \quad \pi(A) = \int \pi(dx)P(x, A)$$

であるような, 状態空間  $E$  に値をとる確率変数列  $X_n$  ( $n \geq 0$ ) である (不変分布  $\pi$  は確率分布であるとき定常分布ともいう. また通常  $E$  は  $p$  次元ユークリッド空間,  $\pi$  は  $\sigma$ -有限な測度  $\mu$  に関して密度をもつと仮定される). 初期値  $X_0$  が与えられたときの  $X_n$  の条件付き分布は,  $P^n$  を推移核  $P$  を  $n$  回繰り返すことと定義すれば

$$P\{X_n \in A | X_0\} = P^n(X_0, A)$$

となる. 不変分布  $\pi$  はすべての可測集合  $A$  について

$$\lim_{n \rightarrow \infty} P^n(x, A) = \pi(A) \quad \text{for } \pi\text{-almost all } x$$

を満たすとき、連鎖の均衡分布とも呼ばれる。不変分布  $\pi$  をもつマルコフ連鎖は初期状態にかかわらず  $\pi$  が正の確率を与える集合にはいる確率が正であるとき  $\pi$ -既約であるといい、また一定の時間間隔で必ず訪れる状態空間のないとき非周期的であるという。マルコフ連鎖  $P$  が  $\pi$ -既約かつ非周期的であり、 $\pi(A) = \int \pi(dx)P(x, A)$  であるならば、 $\pi$  は唯一の不変分布、また連鎖の均衡分布である (例えば Tierney (1994) を参照)。

## 2.2 ギブス・サンプラー

ギブス・サンプラーは、メトロポリス - ヘイスティングス・アルゴリズムの特別な場合であり、その名前は Geman and Geman (1984) が離散分布であるギブス分布から確率標本を得るために用いたアルゴリズムに由来する。その後、データ拡大法 (data augmentation) という欠損値をシミュレートするためのアルゴリズムにも応用されているが、統計学の分野で広く使われるようになったのは Gelfand and Smith (1990) がギブス・サンプラーをベイズ的統計推測の枠組みで広く連続分布にも応用してからである。

いま目標分布  $\pi$  の確率密度関数が  $\pi(\theta)$  であるとし、 $\theta$  は  $\theta = (\theta_1, \dots, \theta_p)$  といくつかのベクトルに分割できるとする。ベイズ推論においては  $\theta$  はパラメータであり、 $\pi(\theta)$  はその事後確率密度関数である。このとき

- (1) 初期値  $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})$  を適当な分布から発生させ、 $i = 0$  とおく。
- (2)  $\theta^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \dots, \theta_p^{(i)})$  が得られたら ( $i \geq 0$ )
  - (i)  $\theta_1^{(i+1)}$  を  $\pi(\theta_1^{(i+1)} | \theta_2^{(i)}, \dots, \theta_p^{(i)})$  から発生させる。
  - (ii)  $\theta_2^{(i+1)}$  を  $\pi(\theta_2^{(i+1)} | \theta_1^{(i+1)}, \theta_3^{(i)}, \dots, \theta_p^{(i)})$  から発生させる。
  - (iii)  $\theta_3^{(i+1)}$  を  $\pi(\theta_3^{(i+1)} | \theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i)}, \dots, \theta_p^{(i)})$  から発生させる。同様に  $\theta_4^{(i+1)}, \dots, \theta_p^{(i+1)}$  を発生させていく。
- (3)  $\theta^{(i+1)} = (\theta_1^{(i+1)}, \theta_2^{(i+1)}, \dots, \theta_p^{(i+1)})$  が得られたら、 $i$  を 1 増やして (2) に戻る。

このとき  $\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \dots$  は

$$\begin{aligned} K(\theta^{(i)}, \theta^{(i+1)}) &= \pi(\theta_1^{(i+1)} | \theta_2^{(i)}, \dots, \theta_p^{(i)}) \\ &\quad \times \prod_{j=2}^{p-1} \pi(\theta_j^{(i+1)} | \theta_1^{(i+1)}, \dots, \theta_{j-1}^{(i+1)}, \theta_{j+1}^{(i)}, \dots, \theta_p^{(i)}) \\ &\quad \times \pi(\theta_p^{(i+1)} | \theta_1^{(i+1)}, \dots, \theta_{p-1}^{(i+1)}) \end{aligned}$$

を推移核とするマルコフ連鎖であり、 $\theta^{(n)}$  の分布は  $n \rightarrow \infty$  のときに  $\pi(\theta)$  を確率密度関数とする分布に収束する。

### 2.3 メトロポリス - ヘイスティングス・アルゴリズム

ギブス・サンプラーでは条件付分布から確率標本を発生することが容易であると仮定したが、実際の応用においては必ずしもこの仮定は成り立たない。その場合には、以下に述べるようなメトロポリス - ヘイスティングス・アルゴリズム (MH アルゴリズム) を行う。

まず不変分布を  $\pi$  とし、 $\mu$  に関して密度関数  $\pi(\theta)$  をもち、退化した分布ではないとする。また推移核  $Q$  を  $Q(\theta, d\theta') = q(\theta, \theta')\mu(d\theta')$ 、 $E^+ = \{\theta : \pi(\theta) > 0\}$  として  $\theta \notin E^+$  は  $Q(\theta, E^+) = 1$  を満たすとする。すると MH アルゴリズムの推移核  $P_{MH}$  は次のように定義される。

$$P_{MH}(\theta, d\theta') = p(\theta, \theta')\mu(d\theta') + r(\theta)\delta_\theta(d\theta'), \quad (1)$$

ただし

$$p(\theta, \theta') = \begin{cases} q(\theta, \theta')\alpha(\theta, \theta'), & \theta \neq \theta' \text{ のとき} \\ 0, & \theta = \theta' \text{ のとき} \end{cases}$$

$$r(\theta) = 1 - \int p(\theta, \theta')\mu(d\theta'),$$

$$\alpha(\theta, \theta') = \begin{cases} \min\left(\frac{\pi(\theta')q(\theta, \theta')}{\pi(\theta)q(\theta', \theta)}, 1\right), & \pi(\theta)q(\theta, \theta') > 0 \text{ のとき} \\ 1, & \pi(\theta)q(\theta, \theta') = 0 \text{ のとき} \end{cases}$$

であり、 $\delta_\theta(d\theta')$  は  $\theta \in d\theta'$  のとき 1、それ以外の場合 0 である関数とする。

具体的には MH アルゴリズムは以下のように進められる。

- (1) 現在の点が  $\theta^{(i)} = \theta$  であるとき提案密度  $q(\theta, \theta')$  を用いて  $\theta'$  を発生させ  $\theta^{(i+1)}$  の候補とする。
- (2) (1) で得られた  $\theta'$  を確率  $\alpha(\theta, \theta')$  で  $\theta^{(i+1)}$  として受容する。棄却した場合には  $\theta^{(i+1)} = \theta$  とする。
- (3) (1) に戻る。

このとき  $\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \dots$  は  $P_{MH}$  を推移核とするマルコフ連鎖であり、 $\theta^{(n)}$  の分布は  $n \rightarrow \infty$  のときに  $\pi(\theta)$  を確率密度関数とする分布に収束する<sup>4</sup>。推移核  $P_{MH}$  によるマルコフ連鎖において、 $\pi(\theta)$  が不変分布の密度関数になるための十分条件は

$$\pi(\theta)p(\theta, \theta') = \pi(\theta')p(\theta', \theta) \quad (2)$$

である。この (2) 式は、均衡状態においては  $\theta$  から  $\theta'$  へ移動する割合と  $\theta'$  から  $\theta$  へ移動する割合は同じであるということの意味しており、詳細釣り合方程式 (detailed balance equation) または可逆性条件 (reversibility condition) と呼ばれている。

ここでは  $\theta$  を一度にサンプリングする方法として説明したが、ギブス・サンプラーにおけるように、 $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p)$  が与えられたときに  $\theta_i$  をサンプリングする際に用いることもできる。逆に言えば、そのような MH アルゴリズムにおいて提案密度  $q(\theta_i, \theta'_i | \theta_{-i})$  が条件付密度  $\pi(\theta'_i | \theta_{-i})$  に一致する場合にギブス・サンプラーとなる。

<sup>4</sup> $P_{MH}$  が  $\pi$ -既約で非周期的であるならば、すべての  $x$  について  $n \rightarrow \infty$  のとき  $\|P_{MH}^n(x, \cdot) - \pi\| \rightarrow 0$  が成り立つ (Tierney (1994))。

## 2.4 収束の判定方法

実際にマルコフ連鎖モンテカルロ法を行う際には、標本が不変分布にいつ収束するのかを判定する必要がある。初期値に依存する期間は稼働検査期間 (burn-in period) であるとして標本を棄て、それ以降の標本を不変分布に収束していると考えて推論を行う<sup>5</sup>。以下では得られた標本系列を用いてマルコフ連鎖が不変分布に収束しているかどうかを検査する方法について説明する<sup>6</sup>。

収束を判定する方法で最も簡単な方法は標本の時系列プロットである。標本経路が初期値に依存せず安定的な動きになっていれば不変分布に収束したと判定すればよい。しかし図による方法では判断が難しい場合も多く、Geweke (1992) は標本系列の平均の安定性に注目して次のような検定方法を提案した。 $\theta^{(t)}$  ( $t = 1, 2, \dots, n$ ) をマルコフ連鎖、その関数  $g(\theta^{(t)})$  を  $g^{(t)}$ ,  $g^{(t)}$  のスペクトル密度関数を  $f(\omega)$  とするとき

- (1)  $\theta^{(t)}$  ( $t = 1, 2, \dots, n$ ) を発生させて、前半  $n_1$  個と後半  $n_2$  個の  $g^{(t)}$  の標本平均をそれぞれ

$$\bar{g}_1 = \frac{1}{n_1} \sum_{t=1}^{n_1} g^{(t)}, \quad \bar{g}_2 = \frac{1}{n_2} \sum_{t=n-n_2+1}^n g^{(t)},$$

とする (Geweke (1992) は経験的に  $n_1 = 0.1n$ ,  $n_2 = 0.5n$  を推奨)。また、それぞれの系列を用いたスペクトル密度の  $\omega = 0$  における推定値を  $\hat{f}_1(0)$ ,  $\hat{f}_2(0)$  とおく。

- (2) 標準正規分布の  $z_\alpha$  を上側  $100\% \alpha$  点とし、

$$Z = \frac{\bar{g}_1 - \bar{g}_2}{\sqrt{2\pi \hat{f}_1(0)/n_1 + 2\pi \hat{f}_2(0)/n_2}}$$

とする。もし、 $|Z| \leq z_{\alpha/2}$  ならば「収束をしていないとはいえない」と判定する<sup>7</sup>。

一方、Gelman (1996), Gelman and Rubin (1992) は、複数のマルコフ連鎖を構成しすべての連鎖が不変分布に収束しているかどうかを、分散の推定値を用いて次のように診断することを提案している。

- (1) 長さ  $2n$  の独立なマルコフ連鎖を  $m$  個発生させて、第  $i$  の連鎖の第  $j$  番目の観測値を  $\theta_i^{(j)}$ , その関数  $g(\theta_i^{(j)})$  を  $g_i^{(j)}$  とおく ( $i = 1, \dots, m, j = 1, \dots, 2n$ )。
- (2) 各連鎖について稼働検査期間を  $n$  とし、後半の  $n$  個の標本を用いて次の  $\sqrt{R}$  を計算する。

$$\sqrt{R} = \sqrt{1 + \frac{1}{n} \left( \frac{B}{W} - 1 \right)}.$$

<sup>5</sup>初期値を定常分布からの確率標本として得る方法に完全シミュレーション (perfect simulation) がある (Propp and Wilson (1996), Murdoch and Green (1998))。

<sup>6</sup>どの方法も完全ではなく (Cowles, Roberts and Rosenthal (1999)) 現実的にはいくつかの方法を併用して収束の判定をする。Robert and Casella (2004), Mengersen, Robert and Guihenneuc-Jouyau (1999) Cowles and Carlin (1996), Brooks and Roberts (1999) を参照。

<sup>7</sup>この方法はスペクトル密度の推定で用いられるウィンドウのとりかたに左右されやすいとの指摘もある (Cowles and Carlin (1996))。



ただし

$$B = \frac{n}{m-1} \sum_{i=1}^m (\bar{g}_i - \bar{\bar{g}})^2, \quad W = \frac{1}{m} \sum_{i=1}^m \frac{\sum_{j=n+1}^{2n} (g_i^{(j)} - \bar{g}_i)^2}{n-1}$$

$$\bar{g}_i = \frac{1}{n} \sum_{j=n+1}^{2n} g_i^{(j)}, \quad \bar{\bar{g}} = \frac{1}{m} \sum_{i=1}^m \bar{g}_i,$$

とする。

- (3)  $\sqrt{R}$  が 1 に近い (1.1 以下,  $R$  では 1.1 ~ 1.2 以下) 場合には収束しているとする。そうでない場合には収束していないとし,  $n$  の値をより大きな値に取り直して再び連鎖を発生させる。

$B/n$  は連鎖間での変動を  $W$  は各連鎖内の変動を表わしており,  $B$  も  $W$  も同じく  $g_i^{(j)}$   $j = 1, 2, \dots$  の分散の推定量であるので, それらが等しいかどうかによって収束の判定を行う。この方法を行う際には, 不変分布の形状に関する情報をできるだけ得て, 初期値が不変分布の台に広く散らばるように発生させて, 状態空間のいろいろな場所から連鎖を始めることが必要である。

事後分布の分位数に関心がある場合には, 幾つかの分位数に注目してマルコフ連鎖が収束しているかどうかを調べる方法もある (Raftery and Lewis (1992), Raftery and Lewis (1996))。Raftery and Lewis (1992) はマルコフ連鎖から  $k$  個おきに標本を抽出した場合について, 一定の推定精度を満たすような  $k$  と初期値に依存する稼働検査期間  $M = mk$  と標本を保存する期間  $N = nk$  を次のように導いている。

マルコフ連鎖  $\theta$  の関数  $U = g(\theta)$  に対して, 累積確率  $q = Pr(U \leq u | \mathbf{y})$  ( $\mathbf{y}$  はデータで,  $u$  は分位数) の点推定量  $\hat{q}$  を

$$\hat{q} = \sum_{t=m+1}^{m+n} Z_{1+(t-1)k}, \quad Z_t = I(U_t \leq u), \quad (I \text{ は定義関数})$$

として  $q$  の  $100(1-\gamma)\%$  信頼区間を  $\hat{q} \pm r$  と求めることを考える (Raftery and Lewis (1992) は  $q = 0.025, \gamma = 0.05, r = 0.0125$  を推奨)。

- (1) まず

$$N = \frac{z_{\gamma/2}^2 q(1-q)}{r^2}$$

個の標本をマルコフ連鎖モンテカルロ法を用いて発生させる。

- (2)  $N$  個の標本から  $\hat{q}$  を求め,  $\alpha = 1 - \hat{q}, \beta = \hat{q}$  とする。

- (3)  $N$  個の標本から, 各  $k$  の値ごとに  $Z_t^{(k)} = Z_{1+(t-1)k}$  ( $t = m+1, \dots, m+n$ ) が 1 次のマルコフ過程か 2 次のマルコフ過程かを BIC 等のモデル選択基準によって選択する。このとき, 1 次のマルコフ過程を選択する最初の  $k$  を求める。

(4)  $m, n$  を

$$m = \frac{\log\left(\frac{\epsilon(\alpha + \beta)}{\max(\alpha, \beta)}\right)}{\log(1 - \alpha - \beta)}, \quad n = \frac{z_{\gamma/2}^2 \alpha \beta (2 - \alpha - \beta)}{r^2 (\alpha + \beta)^3}$$

として,  $M = mk, N = nk$  とする (ここで  $\epsilon$  は初期値に依存する  $Z_t^{(k)}$  の分布とその定常分布の確率関数の差を意味し, Raftery and Lewis (1992) は  $\epsilon = 0.01$  を推奨). 必要ならば (2) で発生させた標本系列に追加してマルコフ連鎖モンテカルロ法を行い, (2) に戻る.

Brooks and Roberts (1999) では分位数のとり方によって収束の速さが異なるため, マルコフ連鎖の収束判定に使用する際には注意が必要であるとしている.

事後分布からのサンプリングを行う場合には, (1) 独立な複数の初期値を  $m$  個発生させて,  $m$  個の初期値それぞれについて稼働検査期間を経たら標本を  $k$  個ずつ保存して  $n = mk$  個得る方法 (多重連鎖, multiple chain) と (2) 1つの初期値から長い1つの系列を発生させ, 稼働検査期間を除いた  $n$  個の標本を使って推論を行う方法 (単一連鎖, single chain) がある. 単一連鎖は状態空間を広く動き回ると期待されるが, 標本の自己相関が高い時には動きは緩慢となり非効率的なサンプリングとなる. 一方, 多重連鎖では得られる標本は独立だが, 初期値を適切に設定し, かつ十分な長さのサンプリングを行わなければ不変分布に収束しない可能性がある.

## 2.5 効率性の評価

マルコフ連鎖の収束判定の後には, サンプリングが効率的に行われているかどうかを調ることにより, どの程度の標本数を用いて推論を行えばよいかを判断する必要がある. サンプリングの効率性を評価するためには, マルコフ連鎖モンテカルロ法によって得られる確率標本の自己相関関数がよく用いられる. 自己相関が高いということは, 標本が現在の場所に留まりがちになり状態空間を万遍なくサンプリングできないことを意味する. 標本の自己相関が低いならば, 動きが独立な確率標本に近くなり状態空間を効率的にサンプリングできることが期待される. したがって効率性の最も簡単な診断方法は, 標本の関数  $g(\theta^n)$  の標本自己相関関数をプロットすること (コレログラムという) である. 縦軸に  $k$  期のラグの標本の自己相関 ( $g(\theta^{(t)})$  と  $g(\theta^{(t+k)})$  の標本相関係数), 横軸に  $k = 1, 2, \dots$ , をプロットしたとき, 標本自己相関関数が急速に減衰していればサンプリングは効率的であるといえる.

また効率性の尺度として非効率性因子 (inefficiency factor) または自己相関時間 (autocorrelation time) がある. マルコフ連鎖  $g(\theta^{(t)})$  の不変分布の分散が存在して  $\sigma^2$  であるとする, その標本平均  $\bar{g}$  の分散は  $(\sigma^2/n)\{1 + 2\sum_{k=1}^n \frac{n-k}{n}\rho(k)\}$  である (ただし  $\rho(k)$  はラグ  $k$  の自己相関関数). もし仮に  $g(\theta^{(t)})$  が互いに独立であるときには, その標本平均  $\bar{g}$  の分散は  $\sigma^2/n$  となるので, この2つの標本平均の分散比  $Var(\bar{g})/\{\sigma^2/n\} = 1 + 2\sum_{k=1}^n \frac{n-k}{n}\rho(k)$  を用い,  $n \rightarrow \infty$  として

$$1 + 2 \sum_{k=1}^{\infty} \rho(k) \quad (3)$$

を非効率性因子 (inefficiency factor) という (Chib (2001)) . この逆数はサンプリングの相対数値的効率性 (relative numerical efficiency)(Geweke (1992)) やサンプリングの漸近的効率性 (Gelman, Roberts and Gilks (1996)) とも呼ばれている . 非効率性因子の値が  $m$  であるとき, 独立な標本を  $n$  個サンプリングした場合と同じ精度の標本平均を得るには,  $mn$  個の標本をマルコフ連鎖モンテカルロ法により発生させる必要がある .

非効率性因子を計算するには, まず (3) 式に推定された標本自己相関係数を代入するか, マルコフ連鎖から得られる標本平均の分散  $Var(\bar{g})$  の推定値を仮想的な独立標本の標本分散の推定値  $s^2/n = \sum_{t=1}^n \{g(\theta^{(t)}) - \bar{g}\}^2 / \{n(n-1)\}$  で割ればよい<sup>8</sup>.

サンプリングの効率性を改善するためには, 提案分布を工夫することが必要である . パラメータ  $\theta$  の次元が大きい場合には,  $\theta$  の成分を 1 つずつ発生させるのではなく, パラメータをいくつかの小さなベクトルに分けて標本を発生させると, サンプリングの効率性が改善される . このような方法はブロック化 (blocking) あるいは多重移動 (multi-move) サンプリングという . その際, 事後分布における相関が高い成分を 1 つのベクトルにまとめるとより効率的なサンプリングとなる .

また, パラメータの変換によってサンプリングする状態空間が変わり, 効率性が改善される場合もある . 分散分析モデルや順序プロビットモデル, 変量効果のあるポアソン分布モデルでは変数変換が有効であることが知られている (Chen, Shao and Ibrahim (2000)).

## 2.6 計算プログラムの検査

マルコフ連鎖を発生するプログラムは複雑になり易いため, 誤りの存在を見つけることが難しい場合も多い . そこで Geweke (2004) はプログラムの正しさをチェックする方法として事後シミュレーション比較 (posterior simulation comparison) を提案している . まず  $y$  の確率密度関数を  $f(y|\theta)$ , パラメータ  $\theta$  の (積分可能な) 事前確率密度関数を  $\pi(\theta)$ , 事後確率密度関数を  $\pi(\theta)$  とおく . 通常事前分布は既知の分布であるので  $\theta \sim \pi(\theta)$  であるような乱数発生は容易であり, モーメントなどに関する情報も既知であることが多い . 一方, マルコフ連鎖を発生するプログラムにデータを生成する部分を付け加えて

- (1)  $\theta \sim \pi(\theta|y)$  を発生する .
- (2)  $y \sim f(y|\theta)$  を発生し, (1) に戻る .

とすることにより得られる  $\theta$  も  $\pi(\theta)$  からの確率標本になっている . そこで  $\theta \sim \pi(\theta)$  という方法と, (1)(2) を反復するという 2 つの方法によって得られる  $\theta$  の標本の 1 次と 2 次のモーメントが (ただし事前分布の 2 次モーメントが存在する場合) 等しいかどうかを仮説検定し, 等しいという帰無仮説が棄却されればプログラムに誤りがあると判断することができる .

<sup>8</sup> , 自己相関関数  $\rho(k)$  が減衰して  $\rho(k) = 0$  ( $k > B_M$ ) とみなせるようなラグについては  $\rho(k) = 0$  とする . 標本平均の分散は, 例えばスペクトル密度関数を用いて  $2\pi\hat{f}(0)/n$  を用いればよい, 推定の際に必要なバンド幅は  $B_M$  とすればよい .

## 2.7 モデル選択

### 2.7.1 周辺尤度と DIC

候補となるモデルが複数ある場合にはモデル選択の基準を用いて最も良いモデルを選ぶことになるが<sup>9</sup>, ベイズ推論においては周辺尤度 (marginal likelihood) やベイズ・ファクター (Bayes factor) がしばしば用いられる. 以下では事前確率密度関数が積分可能である, 正則な事前分布を仮定する<sup>10</sup>. また候補となるモデル  $M$  が  $K$  個 ( $M = 1, \dots, K$ ) あるとし, そのモデルの事前確率を  $\pi(M = i)$  とおく ( $\sum_{i=1}^K \pi(M = i) = 1$ ).  $\mathbf{y}$  を観測値とし,  $\theta_i, f(\mathbf{y}|\theta_i, M = i), \pi(\theta_i|M = i)$  をそれぞれモデル  $M = i$  におけるパラメータ,  $\mathbf{y}$  の確率密度関数, パラメータの事前確率密度関数とすればモデル  $M = i$  におけるデータ  $\mathbf{y}$  の周辺尤度は

$$m(\mathbf{y}|M = i) = \int f(\mathbf{y}|\theta_i, M = i)\pi(\theta_i|M = i)d\theta_i$$

と定義され, モデル  $M = i$  の  $M = j$  に対するベイズ・ファクター  $B_{ij}$  は

$$B_{ij} = \frac{m(\mathbf{y}|M = i)}{m(\mathbf{y}|M = j)} = \frac{\pi(M = i|\mathbf{y})/\pi(M = i)}{\pi(M = j|\mathbf{y})/\pi(M = j)}$$

と定義される. 帰無仮説を  $H_0 : M = i$ , 対立仮説を  $H_1 : M = j$  とするとき, Jeffreys (1961) は,  $H_0$  に反する証拠 (evidence) があるかどうかの尺度として, 表 1 のようにベイズ・ファクターを用いることを提案している<sup>11</sup>.

周辺尤度以外に最近しばしば用いられるモデルの選択基準として, Spiegelhalter *et al.* (2002) によって提案された DIC (Deviance Information Criterion, 偏差情報量基準) がある. DIC は以下のように, はてはまりのよさにモデルの複雑さのペナルティを加えるモデル選択の基準である. モデルのパラメータを  $\theta$ ,  $\theta$  が与えられたときのデータの確率密度関数を  $f(\mathbf{y}|\theta)$  とおくと, あてはまりのよさを評価するためにベイズ偏差 (Bayesian deviance)  $D(\theta)$  を

$$D(\theta) = -2 \log f(\mathbf{y}|\theta) + 2 \log h(\mathbf{y})$$

と定義すると<sup>12</sup>,  $D(\theta)$  はモデルのあてはまりがよいほど小さくなる. 一方, モデルの複雑さを評価するために, 有効なパラメータ数 (effective number of parameters)  $p_D$  を

$$\begin{aligned} p_D &= E_{\pi(\theta|\mathbf{y})} \{D(\theta)\} - D(\theta^*) \\ &= E_{\pi(\theta|\mathbf{y})} \{-2 \log f(\mathbf{y}|\theta)\} + 2 \log f(\mathbf{y}|\theta^*), \quad \theta^* = E_{\pi(\theta|\mathbf{y})}(\theta) \end{aligned}$$

と定義する (ただし  $E_{\pi(\theta|\mathbf{y})}(\cdot)$  は,  $\pi(\theta|\mathbf{y})$  を確率密度関数とする事後分布に関する期待値). ここで Spiegelhalter *et al.* (2002) は平均的なあてはまりの悪さとモデルの複雑さを両方考慮し

<sup>9</sup>モデルを選択する基準にはいろいろあるが, AIC, BIC, GIC については小西・北川 (2004) を参照. ベイズ推論のためのその他の基準については例えば Spiegelhalter *et al.* (2002), Ando (2006).

<sup>10</sup>事前分布が非正則な場合には, データの周辺分布を定義できないので他のモデル選択の基準を用いる. 例えば Kass and Raftery (1995), Berger and Pericchi (1996), Berger and Mortera (1999)などを参照.

<sup>11</sup>Jeffreys (1961) の Tables of  $K$  による.

<sup>12</sup> $h(\mathbf{y})$  はデータ  $\mathbf{y}$  の関数であり,  $D(\theta)$  を基準化するために用いるが, 特にその必要がなければ  $h(\mathbf{y}) = 1$  とおく. 詳細は Spiegelhalter *et al.* (2002) を参照.

$\log_{10} B_{ij}$	$\log B_{ij}$	$B_{ij}$	$H_0$ に反する証拠 (evidence)
0.0 ~	0.00 ~	1.00 ~	$H_0$ を支持する (Null hypothesis supported)
-0.5 ~ 0.0	-1.15 ~ 0.00	0.32 ~ 1.00	$H_0$ に反する証拠があまりあるとはいえない (Evidence against $H_0$ , but not worth more than a bare mention)
-1.0 ~ -0.5	-2.30 ~ -1.15	0.10 ~ 0.32	$H_0$ に反する証拠が十分にある (Evidence against $H_0$ substantial)
-1.5 ~ -1.0	-3.45 ~ -2.30	0.03 ~ 0.10	$H_0$ に反する証拠が強い (Evidence against $H_0$ strong)
-2.0 ~ -1.5	-4.61 ~ -3.45	0.01 ~ 0.03	$H_0$ に反する証拠が非常に強い (Evidence against $H_0$ very strong)
$-\infty$ ~ -2.0	$-\infty$ ~ -4.61	0.00 ~ 0.01	$H_0$ に反する証拠が決定的である (Evidence against $H_0$ decisive)

表 1: ベイズ・ファクターの解釈 (Jeffreys (1961))

て DIC を

$$\begin{aligned} DIC &= E_{\pi(\boldsymbol{\theta}|\mathbf{y})} \{D(\boldsymbol{\theta})\} + p_D \\ &= D(\boldsymbol{\theta}^*) + 2p_D \end{aligned}$$

と定義し, DIC の最も小さいモデルがよいモデルであるとした. DIC はマルコフ連鎖モンテカルロ法によって得られる確率標本によって推定することができる<sup>13</sup>.

周辺尤度の計算方法については, 次の節で重点サンプリング法による方法と周辺尤度の恒等式に基づく方法について説明する.<sup>14</sup>.

### 2.7.2 重点サンプリング法による周辺尤度の推定方法

周辺尤度  $m(\mathbf{y})$  の単純な推定量は, 事前分布  $\pi(\boldsymbol{\theta})$  から独立な確率標本  $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(n)}$  を用いて

$$\hat{m}_{MC}(\mathbf{y}) = \frac{1}{n} \sum_{j=1}^n f(\mathbf{y}|\boldsymbol{\theta}^{(j)}), \quad (4)$$

<sup>13</sup>その標準誤差を求めるには, 例えば DIC を求める計算を  $K$  回繰り返して  $DIC_k$  ( $k = 1, 2, \dots, K$ ) とし, その標本標準偏差を用いればよい.

<sup>14</sup>ラプラス近似を用いた計算方法については小西・北川 (2004) および Raftery (1996) (Laplace-Metropolis 推定量) を, Gelman and Meng (1998) によるパス・サンプリングや Meng and Wong (1996) によるブリッジ・サンプリングで 2 つのモデルの基準化定数の比を求める方法については Chen, Shao and Ibrahim (2000) を参照されたい. また多重積分については伊庭他 (2005) (第 部) による解説や, 関連する Ogata (1989, 1990) も参照.

として得ることができるが、精度の悪いことが知られている (Raftery (1996)). 事前密度  $\pi(\theta_i)$  と尤度関数  $f(\mathbf{y}|\theta)$  の形状が大きく異なっているとき、尤度関数の値が高い  $\theta$  の値からは遠く離れた  $\theta$  ばかりが事前分布から発生されるため、(4) 式の推定値  $\hat{m}_{MC}(\mathbf{y})$  は不安定になりやすい。そこで重点サンプリング法の考え方を用いて、ある確率密度関数  $g(\theta)$  から独立な確率標本  $\theta^{(1)}, \dots, \theta^{(n)}$  を発生させて (ただし  $\pi(\theta) > 0$  である  $\theta$  について  $g(\theta) > 0$  とする)

$$\hat{m}_{IS}(\mathbf{y}) = \frac{1}{n} \sum_{j=1}^n \frac{f(\mathbf{y}|\theta^{(j)})\pi(\theta^{(j)})}{g(\theta^{(j)})} \quad (5)$$

という推定量も考えられるが、推定量の精度を高めるには事後確率密度関数をよく近似する重点関数  $g$  を見つけることが必要である。Gelfand and Dey (1994) は周辺尤度の逆数に注目して事後分布  $\pi(\theta|\mathbf{y})$  から確率標本  $\theta^{(1)}, \dots, \theta^{(n)}$  を発生させて

$$\frac{1}{\hat{m}_{GD}(\mathbf{y})} = \frac{1}{n} \sum_{j=1}^n \frac{g(\theta^{(j)})}{f(\mathbf{y}|\theta^{(j)})\pi(\theta^{(j)})} \quad (6)$$

とすることを提案している。もし重点関数  $g(\theta)$  として  $\pi(\theta)$  を用いれば尤度関数の調和平均

$$\hat{m}_{NR}(\mathbf{y}) = \left[ \frac{1}{n} \sum_{j=1}^n \frac{1}{f(\mathbf{y}|\theta^{(j)})} \right]^{-1}$$

となる (Newton and Raftery (1994)). しかしこの推定量も分母の  $f(\mathbf{y}|\theta^{(j)})$  に 0 に近い値があると推定値が大きく変動するため不安定であり、 $n$  を十分大きくとる必要がある (Raftery (1996)).

### 2.7.3 恒等式に基づく周辺尤度の推定法

すべての  $\theta$  に対して周辺尤度は

$$m(\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\pi(\theta)}{\pi(\theta|\mathbf{y})} \quad (7)$$

をみたしており、両辺の対数をとると

$$\log m(\mathbf{y}) = \log f(\mathbf{y}|\theta) + \log \pi(\theta) - \log \pi(\theta|\mathbf{y}) \quad (8)$$

となり、この恒等式を用いて周辺尤度  $m(\mathbf{y})$  を計算することができる。この式は  $\theta$  の値には依存しないが、実際の計算においては事後密度の推定値が安定的であるような事後平均やモードなどを使う。尤度関数  $f(\mathbf{y}|\theta)$  や事前確率密度  $\pi(\theta)$  は簡単に計算できることが多いので、ここでは事後確率密度  $\pi(\theta|\mathbf{y})$  の推定方法について説明する<sup>15</sup>。推定には MCMC 法によって得られた標本を用いるが、以下ではギブス・サンプラーにより  $\theta = (\theta_1, \dots, \theta_p)$  を  $\theta_i$  ごとにサンプリ

<sup>15</sup>後述するように一般状態空間モデルではモンテカルロ・フィルタによる尤度関数の数値計算が必要となることが多い。

ングする場合に事後密度の推定方法を行う場合について説明する (Chib (1995))<sup>16</sup>.  $\theta = \theta^*$  のとき事後確率密度  $\pi(\theta^*|\mathbf{y})$  は

$$\pi(\theta^*|\mathbf{y}) = \prod_{i=1}^p \pi(\theta_i^*|\mathbf{y}, \psi_{i-1}^*), \quad \psi_{i-1}^* \equiv (\theta_1^*, \dots, \theta_{i-1}^*),$$

(ただし  $\psi_0^*$  は空集合とする) と書くことができるので

$$\log \pi(\theta^*|\mathbf{y}) = \sum_{i=1}^p \log \pi(\theta_i^*|\mathbf{y}, \psi_{i-1}^*), \quad (9)$$

である. ここで

$$\pi(\theta_i^*|\mathbf{y}, \psi_{i-1}^*) = \int \pi(\theta_i^*|\mathbf{y}, \psi_{i-1}^*, \psi^{i+1}) \pi(\psi^{i+1}|\mathbf{y}, \psi_{i-1}^*) d\psi^{i+1}, \quad \psi^{i+1} \equiv (\theta_{i+1}, \dots, \theta_p)$$

であることから, モンテカルロ積分を用いて  $\pi(\theta_i^*|\mathbf{y}, \psi_{i-1}^*)$  を

$$\begin{aligned} \hat{\pi}(\theta_i^*|\mathbf{y}, \psi_{i-1}^*) &= \frac{1}{M} \sum_{m=1}^M \pi(\theta_i^*|\psi_{i-1}^*, \psi^{i+1,(m)}, \mathbf{y}), \\ \psi^{i+1,(m)} &\equiv (\theta_{i+1}^{(m)}, \dots, \theta_p^{(m)}), \\ \theta_k^{(m)} &\sim \pi(\theta_k|\mathbf{y}, \psi_{i-1}^*, \theta_i^*, \theta_{i+1}^{(m)}, \dots, \theta_{k-1}^{(m)}, \theta_{k+1}^{(m-1)}, \dots, \theta_p^{(m-1)}) \end{aligned}$$

とすれば (9) 式より対数事後密度  $\log \pi(\theta^*|\mathbf{y})$  を推定することができる. ここでは  $\theta$  以外に潜在変数が存在しない場合を考えたが, 潜在変数が存在する場合でも同様に計算することができる. 対数事後確率密度  $\log \pi(\theta^*|\mathbf{y})$  の推定値の数値的な標準誤差の求め方については Chib (1995) を参照されたい.

### 3 SV モデル

通常用いられる簡単な SV モデルは, 次の 2 式から構成される.

$$y_t = \exp(h_t/2)\epsilon_t, \quad \epsilon_t \sim i.i.d. \mathcal{N}(0, 1), \quad (10)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t, \quad \eta_t \sim i.i.d. \mathcal{N}(0, \sigma^2). \quad (11)$$

ここで,  $y_t$  は  $t$  期の価格変化率から平均と自己相関を除去したものである. (10) 式は,  $y_t$  をボラティリティと呼ばれる非負の確率変数  $\exp(h_t/2)$  と過去と独立な標準正規分布に従う確率変数  $\epsilon_t$  の積として表している. (11) 式は, ボラティリティの 2 乗の対数値  $h_t$  が次数 1 の自己回帰モデル (AR(1) モデル) に従うものと仮定している.<sup>17</sup> さらに, 誤差項  $\eta_t$  は過去と独立な平均 0, 分散  $\sigma^2$  の正規分布に従い,  $\epsilon_s$  ( $s = 1, \dots, T$ ) とも独立であると仮定する. 株式市場では, 株価

<sup>16</sup>MH アルゴリズム, AR-MH アルゴリズムについては Chib and Jeliazkov (2001, 2005) を参照. 混合分布モデルの場合には修正が必要 (Frühwirth-Schnatter (2004), Frühwirth-Schnatter (2006)). 数値実験による推定法の比較では周辺尤度の恒等式に基づいた方法の精度がよいという報告がある (Han and Carlin (2001)).

<sup>17</sup>(11) 式を 2 次以上の AR モデルや ARMA モデルに拡張するのは容易である.

が上がった日の翌日より下がった日の翌日の方がボラティリティがより上昇する傾向があることが知られており、そうしたボラティリティ変動の非対称性を捉えるためには、 $\epsilon_t$  と  $\eta_t$  の間に相関を導入する必要があるが、そうした拡張については 5.1 節を参照のこと。また、 $h_t$  の初期値  $h_1$  は  $h_t$  の無条件分布である平均 0、分散  $\sigma^2/(1-\phi^2)$  の正規分布に従うものと仮定する。このモデルで推定すべき未知パラメータは  $(\mu, \phi, \sigma^2)$  であり、この内、重要なのは、 $h_t$  に対するショックの持続性を表すパラメータ  $\phi$  である。本稿では、 $h_t$  は定常的であると考え、 $|\phi| < 1$  であると仮定する。<sup>18</sup> その場合、 $\phi$  が 1 に近ければ近いほどボラティリティに対するショックの持続性が高いことになる。資産市場では、ボラティリティが上昇（低下）するとしばらくボラティリティの高い（低い）日が続くことが知られており、こうした現象をボラティリティ・クラスタリング (volatility clustering) と呼ぶ。このことから、ボラティリティに対するショックは持続性が高いことがわかる。実際、SV モデルを推定すると、 $\phi$  の推定値には 1 に近い値が得られるのが常である。<sup>19</sup>

未知パラメータ  $(\mu, \phi, \sigma^2)$  をまとめて  $\theta$ 、 $\mathbf{y} = (y_1, \dots, y_T)$ 、 $\mathbf{h} = (h_1, \dots, h_T)$  と表すと、SV モデルの尤度関数は次のように表される。

$$\begin{aligned} L(\theta) &= f(\mathbf{y}|\theta) \\ &= \int f(\mathbf{y}|\mathbf{h})f(\mathbf{h}|\theta)d\mathbf{h}, \\ &= \int \dots \int \left[ \prod_{t=1}^T f(y_t|h_t) \right] \left[ f(h_1|\theta) \prod_{t=2}^T f(h_t|h_{t-1}, \theta) \right] dh_1 \dots dh_T. \end{aligned}$$

ここで、

$$\begin{aligned} f(y_t|h_t) &= \frac{1}{\sqrt{2\pi \exp(h_t)}} \exp \left[ -\frac{y_t^2}{2 \exp(h_t)} \right], \\ f(h_t|h_{t-1}, \theta) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{\{h_t - \mu - \phi(h_{t-1} - \mu)\}^2}{2\sigma^2} \right], \\ f(h_1|\theta) &= \frac{1}{\sqrt{2\pi\sigma^2/(1-\phi^2)}} \exp \left[ -\frac{(h_1 - \mu)^2}{2\sigma^2/(1-\phi^2)} \right]. \end{aligned}$$

この積分が解析的に解けないため、SV モデルのパラメータは最尤推定することが難しく、最尤法に代わる推定法が必要になる。

<sup>18</sup>  $\mu = 0$ 、 $\phi = 1$  と仮定するモデルもあり、そうしたモデルをランダム・ウォーク SV モデルと呼ぶ。ランダム・ウォーク SV モデルについて詳しくは、Harvey, Ruiz and Shephard (1994) や Ruiz (1994) を参照のこと。また、SV モデルにおいて、 $\phi = 1$  かどうかを検定する方法については、So and Li (1999) や Wright (1999) を参照のこと。

<sup>19</sup> Jacquier, Polson and Rossi (1994) は、それまでの SV モデルを推定した文献をサーベイし、 $\phi$  の推定値には 0.8 から 0.995 までの値が得られているとしている。



## 4 SVモデルのMCMC法を用いたベイズ推定

### 4.1 パラメータのサンプリング

ベイズ推定法では、まず、未知パラメータ  $\theta$  に適当な事前分布  $\pi(\theta)$  を設定する。従来のベイズ推定法は、事前分布をベイズの定理

$$f(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\pi(\theta)}{\int f(\mathbf{y}|\theta)\pi(\theta)d\theta} \quad (12)$$

によってデータ  $\mathbf{y}$  を観測した後の事後分布  $f(\theta|\mathbf{y})$  に更新し、得られた事後分布に基づいてパラメータの値を推定するというものであった。しかし、ベイズの定理 (12) 式の右辺にある  $f(\mathbf{y}|\theta)$  は尤度であり、したがって、SVモデルのように尤度が解析的に求まらないモデルでは、事後分布をベイズの定理を使って解析的に求めることもできない。そうした場合には、何らかの方法によって事後分布から未知パラメータ  $\theta$  の値をサンプリングし、得られた値に基づいてパラメータの値を推定するという方法がとられる。解析的に求まらない未知の事後分布からのサンプリングを可能にしてくれるのが MCMC 法である。

SVモデルの場合、 $\theta = (\mu, \phi, \sigma^2)$  である。そこで、2.2節で説明したギブス・サンプラーを用いて同時事後分布  $f(\mu, \phi, \sigma^2|\mathbf{y})$  からサンプリングするためには、例えば、 $k = 3$ ,  $\theta_1 = \mu$ ,  $\theta_2 = \phi$ ,  $\theta_3 = \sigma^2$  として、

$$f(\mu|\phi, \sigma^2, \mathbf{y}),$$

$$f(\phi|\mu, \sigma^2, \mathbf{y}),$$

$$f(\sigma^2|\mu, \phi, \mathbf{y}),$$

から繰り返しサンプリングすればよいことになる。しかし、残念ながら、これらの条件付事後分布も解析的に求められない。ところが、潜在変数  $\mathbf{h} = (h_1, \dots, h_T)$  も条件に含めた

$$f(\mu|\phi, \sigma^2, \mathbf{h}, \mathbf{y}), \quad (13)$$

$$f(\phi|\mu, \sigma^2, \mathbf{h}, \mathbf{y}), \quad (14)$$

$$f(\sigma^2|\mu, \phi, \mathbf{h}, \mathbf{y}), \quad (15)$$

は解析的に求められる。ここで、 $(\mu, \phi, \sigma^2)$  はSVモデルの (11) 式だけに含まれるパラメータであることに注意しよう。 $\mathbf{y}$  は (10) 式を通じて潜在変数  $\mathbf{h}$  の情報を与えてくれるので、潜在変数  $\mathbf{h}$  の値が未知の場合には、 $(\mu, \phi, \sigma^2)$  の分布を導出するのに  $\mathbf{y}$  と (10) 式が必要となるが、 $\mathbf{h}$  の値が与えられると、 $\mathbf{y}$  も (10) 式も必要なくなり、 $\mathbf{h}$  と (11) 式だけを考えればよいことになる。その場合、(11) 式は  $\mathbf{h}$  を観測値とする単なる AR(1) モデルになるので、 $\mathbf{h}$  を条件に加えた条件付事後分布 (13)–(15) は解析的に求められるのである。また、 $\mathbf{h}$  が与えられると、 $\mathbf{y}$  は必要なくなるので、条件付事後分布 (13)–(15) は、条件から  $\mathbf{y}$  を削除することができる。

事前分布として、通常、 $\mu$  には正規分布、 $\sigma^2$  には逆ガンマ分布が用いられる。<sup>20</sup>

$$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2), \quad \sigma^2 \sim \text{IG}(\nu_0/2, \delta_0/2).$$

<sup>20</sup>  $\sigma^2$  が逆ガンマ分布に従うというのは、逆数  $1/\sigma^2$  がガンマ分布に従うということである。

$\phi$  の事前分布には、ボラティリティが定常であるとの仮定の下、 $|\phi| < 1$  の範囲で切断された切断正規分布、または  $(1 + \phi)/2$  にベータ分布を仮定する。ベータ分布に従う確率変数は 0 から 1 までの値しかとれないので、 $(1 + \phi)/2$  がベータ分布に従うと仮定すると、 $0 < (1 + \phi)/2 < 1$  より、 $|\phi| < 1$  が満たされる。

こうした事前分布の下では、条件付事後分布 (13), (15) はそれぞれ次のように計算される。<sup>21</sup>

$$\mu|\phi, \sigma^2, \mathbf{h}, \sim \mathcal{N}(\mu_1, \sigma_1^2) \quad (16)$$

$$\sigma^2|\mu, \phi, \mathbf{h} \sim \mathcal{IG}(\nu_1/2, \delta_1/2). \quad (17)$$

ただし、

$$\begin{aligned} \sigma_1^2 &= \frac{\sigma_0^2 \sigma^2}{\sigma_0^2 \{(T-1)(1-\phi)^2 + 1 - \phi^2\} + \sigma^2} \\ \mu_1 &= \sigma_1^2 \left\{ \frac{(1-\phi^2)}{\sigma^2} h_1 + \frac{(1-\phi)}{\sigma^2} \sum_{t=1}^{T-1} (h_{t+1} - \phi h_t) + \frac{\mu_0}{\sigma_0^2} \right\} \\ \nu_1 &= T + \nu_0 \\ \delta_1 &= \delta_0 + (h_1 - \mu)^2 (1 - \phi^2) + \sum_{t=1}^{T-1} \{h_{t+1} - \mu - \phi(h_t - \mu)\}^2. \end{aligned}$$

これらの条件付事後分布からは簡単にサンプリングできる。<sup>22</sup>

$\phi$  の条件付事後分布 (14) は次のように計算される。<sup>23</sup>

$$\begin{aligned} \log f(\phi|\mu, \sigma^2, \mathbf{h}) \\ = \text{定数} + \log \varphi(\phi) - \frac{1}{2\sigma^2} \sum_{t=1}^{T-1} \{h_{t+1} - \mu - \phi(h_t - \mu)\}^2, \quad -1 < \phi < 1. \end{aligned} \quad (18)$$

ここで、 $\pi(\phi)$  を  $\phi$  の事前分布とすると、

$$\log \varphi(\phi) = \log \pi(\phi) - \frac{(h_1 - \mu)^2 (1 - \phi^2)}{2\sigma^2} + \frac{1}{2} \log(1 - \phi^2) \quad (19)$$

である。このような特殊な分布からサンプリングする方法に、Acceptance-Rejection (AR) アルゴリズムと 2.3 節で説明した MH アルゴリズムがある。

いま、確率密度関数  $f(x)$  からサンプリングしたいが、特殊な分布であるため直接サンプリングできないものとしよう。ただし、 $f(x)$  は、少なくとも (18) 式の定数項のような基準化定数 (normalizing constant) 以外の部分は解析的に求められ、既知であるものとする。

AR アルゴリズムでは、まず、 $f(x)$  を近似して、かつ、そこから直接サンプリングでき、取り得るすべての  $x$  で優越条件 (dominance condition)  $f(x) \leq cg(x)$  を満たしているような提案密度関数 (proposal density function)  $g(x)$  と正の定数  $c$  を選択する。それを使って以下の

<sup>21</sup>導出については、渡部 (2005) を参照のこと。

<sup>22</sup>(17) のような逆ガンマ分布からサンプリングするには、ガンマ分布からサンプリングして逆数をとればよい。正規分布やガンマ分布といったよく知られた分布からのサンプリングについては、Ripley (1987) を参照のこと。

<sup>23</sup>導出については、渡部 (2005) を参照のこと。

AR アルゴリズムを実行すると,  $f(x)$  から 1 つの値がサンプリングされる.

#### AR アルゴリズム:

- [1] 提案密度関数  $g(x)$  からサンプリングを行い, 得られた値  $x^{(\text{proposal})}$  を使って受容確率  $p$  を次のように計算する.

$$p = \frac{f(x^{(\text{proposal})})}{cg(x^{(\text{proposal})})}$$

- [2] [1] で得られた値  $x^{(\text{proposal})}$  を確率  $p$  で受容し, 確率  $1 - p$  で棄却する<sup>24</sup>. 受容された場合には,  $x = x^{(\text{proposal})}$  とし, 終了. 棄却された場合には [1] へ戻る.

このサンプリングは 1 回前にサンプリングされた値に依存していないので, MCMC 法ではなく, ランダム・サンプリングである. また, この場合, 受容確率  $p = f(x^{(\text{proposal})})/cg(x^{(\text{proposal})})$  が必ず 1 以下であることを保証するため, 優越条件  $f(x) \leq cg(x)$  が必要になる. このアルゴリズムを用いる際に注意すべきことは, 提案密度関数  $g(x)$  と定数  $c$  をモードの近辺で受容確率  $p$  が 1 に近くなるように選ぶことである. そうしないと, [2] で何度も棄却が続いて, サンプリングに時間がかかってしまう. 優越条件を満たしつつ, モードの近辺で受容確率  $p$  が 1 に近くなるように  $g(x)$  や  $c$  を選ぶのは一般的には難しい. 例えば, モードの近辺で受容確率  $p$  が 1 に近くなるように  $c$  を下げると, 分布の裾の部分で優越条件が満たされなくなる可能性がある. これに対して, 以下の MH アルゴリズムでは, 優越条件は必要ない. そこで, 優越条件を満たしつつ, モードの近辺で受容確率  $p$  が 1 に近くなるように  $g(x)$  や  $c$  を選ぶのが難しい場合には MH アルゴリズムが用いられる.

MH アルゴリズムでも, まず,  $f(x)$  を近似して, かつ, 直接サンプリングできるような提案密度関数  $h(x)$  を選択する. ただし, 既に述べたように, その際, AR アルゴリズムのような優越条件は必要ない. また,  $h(x)$  は 1 回前にサンプリングされた値に依存しても構わない. その上で, 適当な初期値  $x_0$  からスタートして, 以下の MH アルゴリズムを実行すれば,  $f(x)$  から  $N$  個の値  $(x_1, \dots, x_N)$  をサンプリングできる.

#### MH アルゴリズム:

- [1]  $n = 1$  とする.

- [2] 提案密度関数  $h(x)$  からサンプリングし, 得られた値  $x^{(\text{proposal})}$  を使って受容確率  $q$  を次のように計算する.

$$q = \min \left[ \frac{f(x^{(\text{proposal})})h(x_{n-1})}{f(x_{n-1})h(x^{(\text{proposal})})}, 1 \right]$$

- [3]  $x^{(\text{proposal})}$  を確率  $q$  で受容し, 確率  $1 - q$  で棄却する. 受容された場合には,  $x_n = x^{(\text{proposal})}$  とする. 棄却された場合には,  $x_n = x_{n-1}$  とする.

<sup>24</sup>[0, 1] の一様分布からサンプリングを行い, 得られた値を  $u$  とし,  $u < p$  であれば受容, そうでなければ棄却すればよい.

[4]  $n < N$  であれば,  $n = n + 1$  として, [2] に戻る.  $n = N$  であれば, 終了する.

これは1回前にサンプリングされた値  $x_{n-1}$  を使ってサンプリングを行っているので, MCMC 法である. MH アルゴリズムを用いる場合にも, AR アルゴリズム同様, 提案密度関数  $h(x)$  をいかに選ぶかが重要になる.<sup>25</sup> このアルゴリズムでは, [3] で棄却されると1回前にサンプリングされた値をそのまま選ぶので, 受容確率  $q$  が低いと, 続けて何度も同じ値をサンプリングしてしまう. そこで, やはり, モード近辺で  $q$  が1に近くなるように提案密度関数  $h(x)$  を選ぶのが望ましい.

条件付事後分布 (18) からサンプリングするためには,  $x = \phi, \log f(x)$  を (18) 式として MH アルゴリズムを実行すればよい. その際の提案密度関数  $h(\phi)$  の選択の方法はいくつか考えられるが, ここでは Chib and Greenberg (1994) で提案されている方法を紹介する. 彼らは, (18) 式で右辺の  $\log \varphi(\phi)$  を無視したものを 定数 +  $\log h(x)$ , すなわち,

$$\begin{aligned} & \log f(\phi|\mu, \sigma^2, \mathbf{h}) \\ & \approx \text{定数} - \frac{1}{2\sigma^2} \sum_{t=1}^{T-1} \{h_{t+1} - \mu - \phi(h_t - \mu)\}^2 \end{aligned} \quad (20)$$

$$= \text{定数} - \frac{\sum_{t=1}^{T-1} (h_t - \mu)^2}{2\sigma^2} \left\{ \phi - \frac{\sum_{t=1}^{T-1} (h_{t+1} - \mu)(h_t - \mu)}{\sum_{t=1}^{T-1} (h_t - \mu)^2} \right\}^2 \quad (21)$$

$$= \text{定数} + \log h(\phi), \quad -1 < \phi < 1, \quad (22)$$

としている.<sup>26</sup> そうすると,  $h(\phi)$  は平均  $\mu_\phi = \sum_{t=1}^{T-1} (h_{t+1} - \mu)(h_t - \mu) / \sum_{t=1}^{T-1} (h_t - \mu)^2$ , 分散  $v_\phi = \sigma^2 / \left\{ \sum_{t=1}^{T-1} (h_t - \mu)^2 \right\}$  の正規分布を  $(-1, 1)$  の範囲を残して切断した切断正規分布になる. この切断正規分布からは簡単にサンプリングできる. まず, 正規分布  $\mathcal{N}(\mu_\phi, v_\phi)$  からサンプリングし,  $|\phi| < 1$  であれば受容し, それ以外であれば棄却すればよい.<sup>27</sup>  $h(\phi)$  をこのように選択すると, 受容確率  $q$  は,  $q = \varphi(x^{(\text{proposal})}) / \varphi(x_{n-1})$  と簡単になる. ただし,  $\varphi(x)$  は (19) 式で定義される.

## 4.2 ボラティリティのサンプリング

SV モデルの未知パラメータ  $(\mu, \phi, \sigma^2)$  の条件付事後分布は, 条件に潜在変数  $h$  を加えると求まり, かつ, そこからサンプリングできることがわかった. そこで, ギブス・サンプラーを適用するには, 潜在変数  $h$  も,  $(\mu, \phi, \sigma^2)$  同様, 未知パラメータとして扱えばよい. ただし, そうすると,  $h$  も条件付事後分布からサンプリングしなければならない.  $h$  は標本の大きさ  $T$  だけあるので, このサンプリングを効率的に行わないと, 膨大な時間がかかって事実上推定不可能になってしまう. そこで, MCMC 法を用いた SV モデルのベイズ推定では,  $h$  をいかに効率良くサンプリングするかが重要になる. これまでに提案されている  $h$  のサンプリング法には single-move sampler, multi-move sampler, mixture sampler の3つがある.

<sup>25</sup>MH アルゴリズムの提案密度関数の選び方について詳しくは, Chib and Greenberg (1995) を参照のこと.

<sup>26</sup>(20)–(22) 式の定数はすべて異なることに注意.

<sup>27</sup>この方法で,  $|\phi| < 1$  の範囲に入る確率が低い場合には, 他の方法を使う必要がある. 他の方法については, 渡部 (2000, p97) を参照のこと.

#### 4.2.1 single-move sampler

SV モデルの MCMC 法を用いたベイズ推定法を提案した Jacquier, Polson and Rossi (1994) は, 各期の潜在変数を,

$$f(h_t | \mu, \phi, \sigma^2, h_1, \dots, h_{t-1}, h_{t+1}, \dots, h_T, \mathbf{y}) \quad (t = 1, \dots, T) \quad (23)$$

から 1 個ずつサンプリングしており, この方法は single-move sampler と呼ばれる. この条件付分布から効率的にサンプリングする方法はいくつか提案されているが,<sup>28</sup> この方法には大きな問題点がある. 3 節で述べたように, ボラティリティに対するショックは持続性が高いので,  $h$  は互いに強い相関がある. このように相関の高い変数を別々にサンプリングすると ギブス・サンプラー の収束の速度が遅いことが知られている. 実際, Shephard and Pitt (1998) は,  $\phi$  が 1 に近い場合にこの方法を用いると収束の速度が遅いことをシミュレーションによって示している.

#### 4.2.2 mixture sampler

SV モデルの (10) 式は状態変数  $h_t$  の非線形関数であるが, 両辺を二乗して対数を取り,  $y_t^* = \log y_t^2$ ,  $\epsilon_t^* = \log \epsilon_t^2$  と定義すると,

$$y_t^* = h_t + \epsilon_t^*$$

となり, 状態変数  $h_t$  の線形関数となる. この  $\epsilon_t^*$  が正規分布であれば線形ガウス状態空間モデルとなり, 状態変数の効率的なサンプリングの方法であるシミュレーション・スムーザ (simulation smoother) を利用することができる (de Jong and Shephard (1995), Durbin and Koopman (2002)). しかし, この変換後の誤差項  $\epsilon_t^*$  は互いに独立に対数カイ二乗分布  $\log \chi_1^2$  ( $\chi_1^2$  は自由度 1 のカイ二乗分布) に従い, その確率密度関数は

$$f(\epsilon_t^*) = \frac{1}{\sqrt{2\pi}} \exp \left\{ \frac{\epsilon_t^* - \exp(\epsilon_t^*)}{2} \right\}, \quad \epsilon_t^* \in R.$$

となる. Kim, Shephard and Chib (1998) はこの対数カイ二乗分布を混合正規分布を用いて

$$g(\epsilon_t^*) = \sum_{j=1}^K p_j f_N(\epsilon_t^* | m_j, v_j^2), \quad \epsilon_t^* \in R, \quad (24)$$

(ただし  $f_N(\epsilon_t^* | m_j, v_j^2)$  は平均  $m_j$ , 分散  $v_j^2$  の正規分布の確率密度関数を表す) のように近似した<sup>29</sup>. さらに潜在変数  $s_t$  を  $K$  個の分布から 1 つの正規分布を選択する確率変数として導入することで,  $s_t$  の条件付の下でモデルが線形ガウス状態空間モデルになることを利用して, 非常に効率的なマルコフ連鎖モンテカルロ法によるサンプリング方法を提案した. Kim, Shephard and Chib (1998) では  $K = 7$  として  $p_j, m_j, v_j^2$  の値を表 2 (左側) のように与えているが, 図 1 の破線に見られるように近似密度がモードの周辺で真の密度 (実線) とやや乖離している. これに対して Kim, Shephard and Chib (1998) の方法を非対称性のある確率的ボラティリティモデルに拡張した Omori *et al.* (2006) では表 2 (右側) のように  $K = 10$  として更に近似精度を高めており, 図 1 では真の密度 (実線) と近似密度 (点線) の差は見えなくなっている.

$j$	Kim <i>et al.</i> (1998)			Omori <i>et al.</i> (2006)		
	$p_j$	$m_j$	$v_j^2$	$p_j$	$m_j$	$v_j^2$
1	0.04395	1.50746	0.16735	0.00609	1.92677	0.11265
2	0.24566	0.52478	0.34023	0.04775	1.34744	0.17788
3	0.34001	-0.65098	0.64009	0.13057	0.73504	0.26768
4	0.25750	-2.35859	1.26261	0.20674	0.02266	0.40611
5	0.10556	-5.24321	2.61369	0.22715	-0.85173	0.62699
6	0.00002	-9.83726	5.17950	0.18842	-1.97278	0.98583
7	0.00730	-11.40039	5.79596	0.12047	-3.46788	1.57469
8				0.05591	-5.55246	2.54498
9				0.01575	-8.68384	4.16591
10				0.00115	-14.65000	7.33342

表 2:  $(p_j, m_j, v_j^2, a_j, b_j)$  の選択.

潜在変数  $s_t$  を用いて混合正規分布にしたがう確率密度関数を表現すると,  $\Pr(s_t = i) \times f_N(\epsilon_t^* | m_i, v_i^2)$  となるので,  $s_t$  の条件付事後分布は

$$\pi(s_t = i | h, \mu, \theta, \mathbf{y}^*) \propto \Pr(s_t = i) v_i^{-1} \exp \left\{ -\frac{(y_t^* - h_t - m_i)^2}{2v_i^2} \right\}, \quad i = 1, 2, \dots, K,$$

(ただし  $\mathbf{y}^* = (y_1^*, \dots, y_T^*)'$ ) となる. したがって  $s_t$  のサンプリングはこの離散分布から行うことができる.

一方,  $s_t$  を所与とすればモデルは線形ガウス状態空間モデル

$$\begin{aligned} y_t^* &= m_{s_t} + h_t + u_t, \quad u_t \sim \mathcal{N}(0, v_{s_t}^2), \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

となるので,  $\mu, \phi, \sigma^2, \mathbf{s} = (s_1, \dots, s_T)$  を所与として  $\mathbf{h} = (h_1, \dots, h_T)$  の事後分布から同時にサンプリングすることができる. ここでは de Jong and Shephard (1995) によるシミュレーション・スムーザを用いたサンプリングの方法について説明する (ただし  $h_t$  が多次元の場合には Durbin and Koopman (2002) を用いるとよい).

(1) 以下のカルマンフィルタを行い,  $\{e_t, D_t, J_t, L_t\}_{t=1}^T$  を保存する.

(a)  $a_1 = \mu, P_1 = \sigma^2 / (1 - \phi^2)$  とする.

(b)  $t = 1, 2, \dots, T - 1$  のとき

$$a_{t+1} = (1 - \phi)\mu + \phi a_t + K_t e_t, \quad P_{t+1} = \phi P_t L_t + \sigma^2, \quad (25)$$

を計算する. ただし

$$\begin{aligned} e_t &= y_t^* - m_{s_t} - a_t, \quad D_t = P_t + v_{s_t}^2, \quad K_t = \phi P_t / D_t, \\ L_t &= \phi - K_t, \quad J_t = (0, \sigma) - K_t(v_{s_t}, 0). \end{aligned}$$

<sup>28</sup>Shephard and Pitt (1997), 渡部 (2000, pp.98–100) を参照のこと.

<sup>29</sup>近似の方法については Kim, Shephard and Chib (1998) 参照

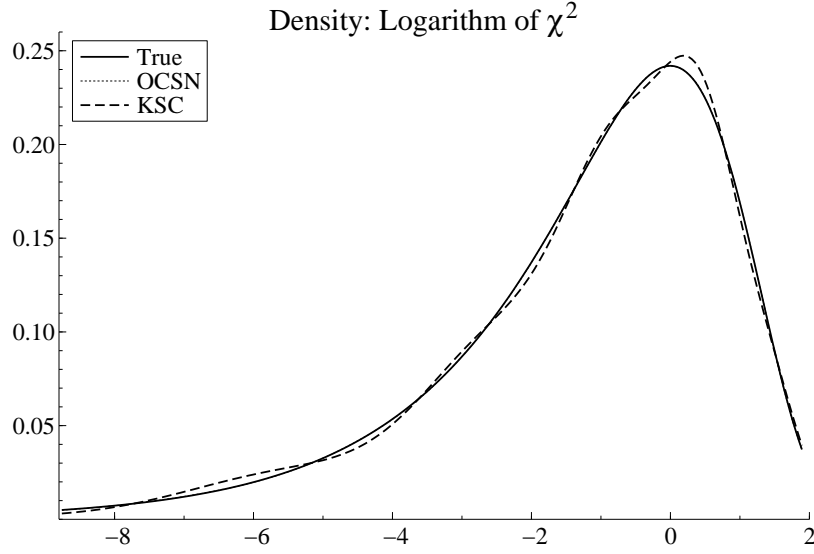


図 1: 対数カイ二乗分布の近似.

実線: 真の密度関数, 点線: Omori *et al* (2006) による近似, 破線: Kim *et al.* (1998) による近似.

(2) 以下のシミュレーション・スムーザを行い,  $\{\xi_t\}_{t=0}^{T-1}$  を保存する.

(a)  $r_T = 0$  と  $U_T = 0$  として,  $t = T, T-1, \dots, 1$  の順に

$$\begin{aligned} C_t &= \sigma^2 - \sigma^4 U_t, \quad \kappa_t \sim \mathcal{N}(0, C_t), \quad V_t = \sigma^2 U_t L_t, \\ r_{t-1} &= D_t^{-1} e_t + L_t r_t - V_t C_t^{-1} \kappa_t, \\ U_{t-1} &= D_t^{-1} + U_t L_t^2 + C_t^{-1} V_t^2, \end{aligned}$$

を逐次的に計算して行き,  $\{\kappa_t\}_{t=1}^T, \{r_t\}_{t=0}^T$  を求める.

(b) さらに  $\kappa_0$  を

$$C_0 = \frac{\sigma^2}{1 - \phi^2} - \frac{\sigma^4}{(1 - \phi^2)^2} U_0, \quad \kappa_0 \sim \mathcal{N}(0, C_0),$$

と求める.

(c) (a)(b) で得られた  $\{\kappa_t\}_{t=0}^{T-1}, \{r_t\}_{t=0}^{T-1}$  より

$$\begin{aligned} \xi_t &= \sigma^2 r_t + \kappa_t, \quad t = 1, \dots, T-1, \\ \xi_0 &= \frac{\sigma^2}{1 - \phi^2} r_0 + \kappa_0, \end{aligned}$$

を求める. このとき  $\xi = (\xi_0, \xi_1, \dots, \xi_{T-1})$  は  $\eta = (\eta_0, \eta_1, \dots, \eta_{T-1})$  の事後分布からの確率標本となる.

(3)  $\xi$  を用いて  $h = (h_1, \dots, h_T)$  の事後分布からの確率標本を

$$\begin{aligned} h_{t+1} &= (1 - \phi)\mu + \phi h_t + \xi_t, \quad t = 1, \dots, T - 1, \\ h_1 &= \mu + \xi_0, \end{aligned}$$

とすればよい。

$(\phi, \sigma^2, \mu)$  のサンプリングについてはすでに説明した通りであるが、混合正規分布による近似を行った場合には、 $(\phi, \sigma^2, s, \mathbf{y}^*)$  を所与として  $(h, \mu)$  を積分した尤度を求めることができる。この尤度と事前確率密度から MH アルゴリズムで (1)  $\phi, \sigma^2 \sim \pi(\phi, \sigma^2 | \mathbf{y}^*, s)$  (2)  $\mu \sim \pi(\mu | \phi, \sigma^2, \mathbf{y}^*, s)$  の順に発生することにより、更に効率的なサンプリングを行うこともできる。このサンプリング方法は de Jong (1991) の拡大カルマン・フィルタ (augmented Kalman filter) を用いた integration sampler と呼ばれる方法で、後述の非対称性を考慮した SV モデルにおいて説明する。

また実際に株式の収益率  $y_t$  (単位はパーセント) を用いて  $y_t^* = \log y_t^2$  を計算する際には、 $y_t \approx 0$  の時に  $y_t^*$  の値が不安定になることから、Kim, Shephard and Chib (1998) では  $y_t^* = \log(y_t^2 + c)$ ,  $c = 0.001$  として計算している。  $c$  のとり方が結果に与える影響には精査が必要であり、 $c$  を更に小さくするためには混合正規分布による近似も改善することが必要となる。Omori *et al.* (2006) では近似を改善したため、シミュレーションデータによる結果では  $c = 0.0001$  としても影響はほとんどないことを確認している。

ただし、mixture sampler にはいくつかの問題点がある。まず、 $z_t$  の正しい分布ではなく、混合正規分布で近似したものを使っているので、あくまでも近似的な方法にすぎないという点である。次に、この方法を適用できるのは、線形状態空間モデルで表現できるモデルだけという点である。例えば、リスク・プレミアムを考慮して、(10) 式を、

$$y_t = a + b \exp(h_t) + \exp(h_t/2)\epsilon_t$$

に置き換えたモデルは、線形状態空間モデルに書き換えることができない。SV モデルを取引高を含める形で発展させたモデルに動学的 2 変量分布混合モデル (5.4 節参照) があるが、それも線形状態空間モデルでは表せない。mixture sampler は、そうしたモデルには適用できない。モデルの拡張性という観点からは、線形状態空間モデルへの変換を必要としない次の multi-move sampler の方が望ましい。

#### 4.2.3 multi-move sampler

(24) 式のような変換を行わず、(10), (11) 式を使って、条件付事後分布  $f(h | \mu, \phi, \sigma^2, \mathbf{y})$  から潜在変数  $h$  を一度にサンプリングするのは、特に  $T$  が大きい場合には難しい。そこで、 $h$  を一度にサンプリングするのではなく、いくつかのブロックに分けて、1つのブロックを一度にサンプリングするという方法を考えよう。例えば、 $(h_t, \dots, h_{t+k})$  が1つのブロックだとすると、それらを、

$$f(h_t, \dots, h_{t+k} | \mu, \phi, \sigma^2, h_{t-1}, h_{t+k+1}, y_t, \dots, y_{t+k}) \quad (26)$$



から一度にサンプリングするということである。<sup>30</sup> しかし、 $(h_t, \dots, h_{t+k})$  には互いに相関があるので、(26) からのサンプリングも容易ではない。そこで、Shephard and Pitt (1997) は、潜在変数  $(h_t, \dots, h_{t+k})$  ではなく、(11) 式の誤差項  $(\eta_{t-1}, \dots, \eta_{t+k-1})$  を、

$$f(\eta_{t-1}, \dots, \eta_{t+k-1} | \mu, \phi, \sigma^2, h_{t-1}, h_{t+k+1}, y_t, \dots, y_{t+k}) \quad (27)$$

からサンプリングするという方法を提案している。 $(\mu, \phi, \sigma^2)$  と  $h_{t-1}$  の値が与えられた下で、 $(\eta_{t-1}, \dots, \eta_{t+k-1})$  がサンプリングされると、(11) 式から、 $(h_t, \dots, h_{t+k})$  を逐次的に計算できる。この方法は、multi-move sampler もしくは block sampler と呼ばれる。

条件付分布 (27) からサンプリングする方法として、Shephard and Pitt (1997) は、Tierney (1994) の提案した Acceptance-Rejection Metropolis-Hastings (ARMH) アルゴリズムを用いている。<sup>31</sup> このアルゴリズムは、既に説明した MH アルゴリズムの候補  $x^{(\text{proposal})}$  を AR アルゴリズムを用いてサンプリングするという方法である。そうすることで、MH アルゴリズムの提案密度関数が正しい密度関数  $f(x)$  をよりうまく近似するようになるので、MH アルゴリズムの受容確率が高まる。このアルゴリズムでも、MH アルゴリズム同様、AR アルゴリズムのような優越条件は必要ない。まず、AR アルゴリズムの提案密度関数  $g(x)$  と正の定数  $c$  を選択し、適当な初期値  $x_0$  からスタートして、以下のアルゴリズムを実行すれば、 $f(x)$  から  $N$  個の値  $(x_1, \dots, x_N)$  をサンプリングできる。

#### ARMH アルゴリズム:

- [1]  $n = 1$  とする。
- [2] 提案密度関数  $g(x)$  からサンプリングを行い、得られた値  $x$  を使って受容確率  $p$  を次のように計算する。

$$p = \min \left[ \frac{f(x)}{cg(x)}, 1 \right]$$

- [3] [2] で得られた値  $x$  を確率  $p$  で受容し、確率  $1 - p$  で棄却する。受容された場合には、 $x^{(\text{proposal})} = x$  とおき、[4] に進む。棄却された場合には [2] へ戻る。
- [4] 受容確率  $q$  を以下のように計算する。

- (a)  $f(x_{n-1}) < cg(x_{n-1})$  ならば、 $q = 1$ ,
- (b)  $f(x_{n-1}) \geq cg(x_{n-1})$  かつ  $f(x^{(\text{proposal})}) < cg(x^{(\text{proposal})})$  ならば、

$$q = \frac{cg(x_{n-1})}{f(x_{n-1})},$$

<sup>30</sup> ここで、 $(h_1, \dots, h_{t-2})$ ,  $(h_{t+k+2}, \dots, h_T)$ ,  $(y_1, \dots, y_{t-1})$ ,  $(y_{t+k+1}, \dots, y_T)$  は条件から削除されていることに注意されたい。(11) 式より、 $h_t$  は 1 期前の値にのみ依存するので、 $h_{t-1}$  と  $h_{t+k+1}$  が与えられると、 $(h_t, \dots, h_{t+k})$  の分布を導出するのに、 $(h_1, \dots, h_{t-2})$  と  $(h_{t+k+2}, \dots, h_T)$  は必要ない。また、 $(y_1, \dots, y_{t-1})$ ,  $(y_{t+k+1}, \dots, y_T)$  がもたらすのは  $(h_1, \dots, h_{t-1})$  と  $(h_{t+k+1}, \dots, h_T)$  の情報だけなので、それらも必要ない。

<sup>31</sup> ARMH アルゴリズムについて詳しくは、Chib and Greenberg (1995)、渡部 (2000 3.5.3 節) 参照。

(c)  $f(\mathbf{x}_{n-1}) \geq cg(\mathbf{x}_{n-1})$  かつ  $f(\mathbf{x}^{(\text{proposal})}) \geq cg(\mathbf{x}^{(\text{proposal})})$  ならば,

$$q = \min \left[ \frac{f(\mathbf{x}^{(\text{proposal})})g(\mathbf{x}_{n-1})}{f(\mathbf{x}_{n-1})g(\mathbf{x}^{(\text{proposal})})}, 1 \right].$$

[5]  $\mathbf{x}^{(\text{proposal})}$  を確率  $q$  で受容し, 確率  $1-q$  で棄却する. 受容された場合には,  $\mathbf{x}_n = \mathbf{x}^{(\text{proposal})}$  とする. 棄却された場合には,  $\mathbf{x}_n = \mathbf{x}_{n-1}$  とする.

[6]  $n < N$  であれば,  $n = n + 1$  として (2) に戻る.  $n = N$  であれば, 終了.

この内, [2], [3] が AR アルゴリズム, [4], [5] が MH アルゴリズムである. このアルゴリズムでは, 優越条件  $f(\mathbf{x}) \leq cg(\mathbf{x})$  は必要はないが, もし,  $\mathbf{x}$  の取り得るすべての値でそれが満たされるならば, MH パートの受容確率は常に  $q = 1$  となり, AR アルゴリズムでサンプリングされた値を常に受容するので, ARMH アルゴリズムは通常の AR アルゴリズムになる.

条件付分布 (27) から  $(\eta_{t-1}, \dots, \eta_{t+k-1})$  をサンプリングするためには,  $\mathbf{x} = (\eta_{t-1}, \dots, \eta_{t+k-1})$ , (27) を  $f(\mathbf{x})$  として, ARMH アルゴリズムを実行すればよい. その際, 提案密度関数  $g(\mathbf{x})$  と定数  $c$  を選択しなければならないが, 効率的にサンプリングを行うためには,  $g(\mathbf{x})$  および定数  $c$  を AR パートおよび MH パートの受容確率  $p, q$  ができるだけ 1 に近くなるように選ぶのが望ましい. そのため, Shephard and Pitt (1997) は,  $g(\mathbf{x})$  を条件付事後分布 (27) をうまく近似するような  $k + 1$  変量正規分布として選んでいる. 以下, 彼らの  $g(\mathbf{x})$  の選択の仕方およびそこからのサンプリングの方法について説明しよう.

まず,  $t + k < T$  の場合を考えよう. この場合, 条件付事後分布 (27) の対数をとったものは次のように表せる.<sup>32</sup>

$$\begin{aligned} & \log f(\eta_{t-1}, \dots, \eta_{t+k-1} | \mu, \phi, \sigma^2, h_{t-1}, h_{t+k+1}, y_t, \dots, y_{t+k}) \\ &= \text{定数} + \sum_{s=t}^{t+k} \log f(y_s | h_s) + \log f(h_{t+k+1} | \mu, \phi, \sigma^2, h_{t+k}) + \sum_{s=t-1}^{t+k-1} \log f(\eta_s | \sigma^2), \\ &= \text{定数} - \sum_{s=t}^{t+k} \left\{ \frac{h_s}{2} + \frac{y_s^2}{2} \exp(-h_s) \right\} - \frac{1}{2\sigma^2} \{h_{t+k+1} - \mu - \phi(h_{t+k} - \mu)\}^2 \\ & \quad - \frac{1}{2\sigma^2} \sum_{s=t-1}^{t+k-1} \eta_s^2. \end{aligned} \tag{28}$$

Shephard and Pitt (1997) は,

$$l(h_s) = -\frac{h_s}{2} - \frac{y_s^2}{2} \exp(-h_s) \tag{29}$$

を  $\hat{h}_s$  の回りで 2 次までテーラー展開することにより (28) 式を以下のように近似したものを

<sup>32</sup>導出については, 渡部 (2005) を参照のこと.

$\log(cg(\eta_{t-1}, \dots, \eta_{t+k-1}))$  としている.

$$\begin{aligned}
& \log f(\eta_{t-1}, \dots, \eta_{t+k-1} | \mu, \phi, \sigma^2, h_{t-1}, h_{t+k+1}, y_t, \dots, y_{t+k}) \\
& \approx \text{定数} + \sum_{s=t}^{t+k} \left\{ l(\hat{h}_s) + (h_s - \hat{h}_s)l'(\hat{h}_s) + \frac{1}{2}(h_s - \hat{h}_s)^2 l''(\hat{h}_s) \right\} \\
& \quad - \frac{1}{2\sigma^2} \{h_{t+k+1} - \mu - \phi(h_{t+k} - \mu)\}^2 - \frac{1}{2\sigma^2} \sum_{s=t-1}^{t+k-1} \eta_s^2, \\
& = \text{定数} - \sum_{s=t}^{t+k-1} \frac{1}{2v_s} (\hat{y}_s - h_s)^2 - \frac{1}{2\sigma^2} \sum_{s=t-1}^{t+k-1} \eta_s^2, \\
& = \log(cg(\eta_{t-1}, \dots, \eta_{t+k-1})). \tag{30}
\end{aligned}$$

ここで,  $v_s, \hat{y}_s$  は以下のように定義される.  $s = t, \dots, t+k-1$  であれば,

$$v_s = -\frac{1}{l''(\hat{h}_s)}, \tag{31}$$

$$\hat{y}_s = \hat{h}_s + v_s l'(\hat{h}_s), \tag{32}$$

$s = t+k < T$  であれば,

$$v_s = \frac{\sigma^2}{\phi^2 - \sigma^2 l''(\hat{h}_s)}, \tag{33}$$

$$\hat{y}_s = \hat{h}_s + v_s \left[ l'(\hat{h}_s) + \frac{\phi}{\sigma^2} \{h_{t+k+1} - \mu - \phi(\hat{h}_s - \mu)\} \right]. \tag{34}$$

ただし,

$$l'(\hat{h}_s) = \frac{1}{2} \{y_s^2 \exp(-\hat{h}_s) - 1\} \tag{35}$$

$$l''(\hat{h}_s) = -\frac{y_s^2}{2} \exp(-\hat{h}_s) \tag{36}$$

である.

次に,  $t+K = T$  の場合, すなわち, 最後のブロックを考えよう. その場合, 条件に  $h_{t+k+1}$  は含まれないので, (28) 式を次のように書き換えなければならない.

$$\begin{aligned}
& \log f(\eta_{t-1}, \dots, \eta_{t+k-1} | \mu, \phi, \sigma^2, h_{t-1}, y_t, \dots, y_{t+k}) \\
& = \text{定数} - \sum_{s=t}^{t+k} \left\{ \frac{h_s}{2} + \frac{y_s^2}{2} \exp(-h_s) \right\} - \frac{1}{2\sigma^2} \sum_{s=t-1}^{t+k-1} \eta_s^2. \tag{37}
\end{aligned}$$

そこで, この場合には,  $s = t, \dots, t+k$  すべてで,  $v_s, \hat{y}_s$  は (31), (32) 式で定義される.

以上をまとめると,  $v_s, \hat{y}_s$  は,  $s = t, \dots, t+k-1$  あるいは  $s = t+k = T$  であれば (31), (32) 式,  $s = t+k < T$  であれば (33), (34) 式で定義される. ここで, このようにして定義される  $\hat{y}_s$  を観測値,  $h_s$  を状態変数とする線形ガウス状態空間モデル

$$\hat{y}_s = h_s + \xi_s, \quad \xi_s \sim i.i.d. \mathcal{N}(0, v_s), \tag{38}$$

$$h_s = \mu + \phi(h_{s-1} - \mu) + \eta_s, \quad \eta_s \sim i.i.d. \mathcal{N}(0, \sigma^2) \tag{39}$$

を考えよう.<sup>33</sup> 提案密度関数  $g(\eta_{t-1}, \dots, \eta_{t+k-1})$  はこの線形ガウス状態空間モデルにおける条件付事後分布

$$f(\eta_{t-1}, \dots, \eta_{t+k-1} | \mu, \phi, \sigma^2, h_{t-1}, h_{t+k+1}, y_t, \dots, y_{t+k}) \quad (40)$$

である.<sup>34</sup> また, そこからサンプリングするためには, de Jong and Shephard (1995), Durbin and Koopman (2002) らによつての提案されている simulation smoother を使えばよい.

テーラー展開を行う点  $(\hat{h}_t, \dots, \hat{h}_{t+k})$  は, (28) 式のモードとするのが望ましい. そうすれば, モードの周辺で,

$$f(\eta_{t-1}, \dots, \eta_{t+k-1} | \mu, \phi, \sigma^2, h_{t-1}, h_{t+k+1}, y_t, \dots, y_{t+k}) \approx cg(\eta_{t-1}, \dots, \eta_{t+k-1})$$

となり, ARMH アルゴリズムの受容確率  $p, q$  がいずれもモードの周辺で 1 に近くなる. そのためには,  $(\hat{h}_t, \dots, \hat{h}_{t+k})$  を次のように選べばよい.  $(\hat{h}_t, \dots, \hat{h}_{t+k})$  に適当な初期値を選んでやると, (32), (34) 式より  $(\hat{y}_t, \dots, \hat{y}_{t+k})$  が計算できる. そこで, それを使って, (38), (39) 式から成る線形ガウス状態空間モデルに対してカルマン・フィルタとスムーザを実行すると,  $E(h_{t+i} | \mu, \phi, \sigma^2, h_{t-1}, h_{t+k+1}, y_t, \dots, y_{t+k})$  ( $i = 0, \dots, k$ ) が求まる. それを (32), (34) 式の  $(\hat{h}_t, \dots, \hat{h}_{t+k})$  に代入すると新たな  $(\hat{y}_t, \dots, \hat{y}_{t+k})$  が求まる. 今度は, それを使い, (38), (39) 式から成る線形状態空間モデルにおいて再びカルマン・フィルタとスムーザを実行すると, 新たな  $E(h_{t+i} | \mu, \phi, \sigma^2, h_{t-1}, h_{t+k+1}, y_t, \dots, y_{t+k})$  ( $i = 0, \dots, k$ ) が求まる. これを数回繰り返すと, モードに近い  $(\hat{h}_t, \dots, \hat{h}_{t+k})$  が得られるので, そこでテイラー展開を行えばよい.

このアルゴリズムでは,  $(\eta_1, \dots, \eta_T)$  をいくつかのブロックに分割する必要がある.  $k_0 = 0$ ,  $k_{K+1} = T$  として,  $K + 1$  個のブロック  $(\eta_{k_{i-1}+1}, \dots, \eta_{k_i})$  ( $i = 1, \dots, K + 1$ ) に分割するものとしよう. Shephard and Pitt (1997) は,  $(k_1, \dots, k_K)$  をランダムに選んでいる. 具体的には,  $U_i$  を  $[0, 1]$  の一様分布からサンプリングし,

$$k_i = \text{int}[T \times \{(i + U_i)/(K + 2)\}], \quad i = 1, \dots, K$$

としている. ここで,  $\text{int}[x]$  は,  $x$  に最も近い整数値を表している. このように, ブロックをランダムに選ぶと,  $n$  回目のサンプリングで高い確率で棄却されるブロックがあったとしても,  $n + 1$  回目のサンプリングでは, 異なるブロックが選ばれるので, 棄却が続いてサンプリングが行き詰まってしまうということを排除できる.

以上が multi-move sampler と呼ばれるアルゴリズムであるが, それを提案した Shephard and Pitt (1997) は, (28) 式の

$$\log f(h_{t+k+1} | \mu, \phi, \sigma^2, h_{t+k}) = -\frac{1}{2\sigma^2} \{h_{t+k+1} - \mu - \phi(h_{t+k} - \mu)\}^2$$

の項を無視している. その結果,  $v_s, \hat{y}_s$  を,  $s = t + K < T$  の場合にも, (31), (32) 式で定義している. この点を修正したのが, Watanabe and Omori (2004a,b) であり, Watanabe and Omori (2004a) では, さらに, Shephard and Pitt (1997) の方法をそのまま用いると, パラメータや潜在変数の推定値にバイアスが生じることを示している.

<sup>33</sup>観測方程式 (38) の分散  $v_s$  は (31) または (33) 式で定義されるが, (31), (33) 式の分母の  $l''(\hat{h}_s)$  は (36) 式より非正なので,  $v_s$  は非負である. ただし, (36) 式より,  $y_s = 0$  の場合には,  $l''(\hat{h}_s) = 0$  なので, (31) 式より,  $v_s = \infty$  となってしまう. そこで, 0 または 0 に近い値を含んでいるようなデータの場合には, (31) 式の分母および (30) の  $l''(\hat{h}_s)$  を,  $c$  を 0 に近い負の定数 (例えば,  $-0.00001$ ) として,  $\min[l''(\hat{h}_s), c]$  に置き換えればよい.

<sup>34</sup>(38), (39) 式から成る線形ガウス状態空間モデルの下で, 条件付事後分布 (40) を, (28) 式と同様に展開すれば, (30) が得られる.

## 5 SVモデルの発展

MCMC法を用いたベイズ推定法はモデルを拡張しても適用可能なので、最近ではSVモデルも、ARCH型モデル同様、さまざまな形で拡張が行われるようになってきている。本節では、そうしたSVモデルの発展についてサーベイを行う。

### 5.1 非対称SVモデル

株式市場では、株価が上がった日の翌日より下がった日の翌日の方がボラティリティが高まる傾向があることが知られている (Black (1976), Nelson(1991))<sup>35</sup>。この現象は非対称性 (asymmetry) 或いはレバレッジ効果 (leverage effect) と呼ばれており、これをモデルに取り込むことは株式収益率のリスクを評価する際にも重要である。SVモデルでこうしたボラティリティ変動の非対称性を捉えるためには、以下のように誤差項  $\epsilon_t$  と  $\eta_t$  の間に相関を導入すればよい。

$$y_t = \epsilon_t \exp(h_t/2), \quad t = 1, \dots, T, \quad (41)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t, \quad t = 1, \dots, T-1, \quad (42)$$

$$h_1 \sim \mathcal{N}(0, \sigma^2/(1 - \phi^2)),$$

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim i.i.d. \mathcal{N}(0, \Sigma), \quad \text{ただし } \Sigma = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}. \quad (43)$$

ここで、もし  $\rho < 0$  であれば、株式市場で観測されるボラティリティ変動の非対称性と整合的であり、 $\rho = 0$  であれば、通常非対称性の無いSVモデルになる。

Yu (2005), Omori and Watanabe (2003), Omori *et al.* (2006), Nakajima and Omori (2006) らは、こうした非対称SVモデルをMCMC法を用いてベイズ推定している。Yu (2002) はボラティリティのサンプリングに single-move sampler を用いており、この方法は既に述べたように効率的でない。前節で説明した multi-move sampler は非対称SVモデルには適用できないので、Omori and Watanabe (2003) は、非対称SVモデルの multi-move sampler を新たに提案している<sup>36</sup>。また、Omori *et al.* (2006) は mixture sampler を拡張している。Jacquier, Polson and Rossi (2004) は、 $\epsilon_t$  と  $\eta_t$  の間ではなく、 $\epsilon_t$  と  $\eta_{t-1}$  の間に相関を導入したモデルを single-move sampler を用いてベイズ推定しているが、Yu (2005) は  $\epsilon_t$  と  $\eta_t$  の間に相関を導入した方が当てはまりが良いとの結果を得ている。

非対称SVモデルを他の推定法を用いて推定しているものには、Melino and Turnbull (1990) と Harvey and Shephard (1996) がある。Melino and Turnbull (1990) は一般化積率法 (generalized method of moments; GMM), Harvey and Shephard (1996) はカルマン・フィルタに基づく疑似最尤法 (quasi-maximum likelihood estimation; QML) を用いている。<sup>37</sup> GMM や QML や MCMC法を用いたベイズ推定法と比べて推定量の効率性が低い (Jacquier, Polson and Rossi 1994)。

<sup>35</sup>なぜそのような非対称性が生じるのかを分析しているものに、Christie (1982), Wu (2001) がある。

<sup>36</sup>この方法は mixture sampler では推定できない様々な拡張モデルに適用することが可能であるが、ここでは紙数の制約のため紹介を割愛することとする。

<sup>37</sup>SVモデルのQML推定については、渡部 (2000 3.4節) を参照のこと。

### 5.1.1 mixture sampler

以下では Omori *et al.* (2006) の 2 次元混合正規分布による近似を用いて mixture sampler について説明する．(41)–(42) 式のモデルでは，観測方程式 (41) がボラティリティを表現する潜在変数  $h_t$  の非線形関数になっているが，Kim, Shephard and Chib (1998) のように両辺を二乗して対数をとれば線形関数となり推定することが容易になる．ただし，被説明変数  $y_t$  の正負に関する符号情報は二乗することにより失われてしまうので，次のように変数の変換を考えることにする．

$$d_t = I(\epsilon_t \geq 0) - I(\epsilon_t < 0), \quad (44)$$

$$y_t^* = \log y_t^2 = h_t + \epsilon_t^*, \quad \epsilon_t^* = \log \epsilon_t^2, \quad (45)$$

この 2 変数  $(d_t, y_t^*)$  を用いれば，元の説明変数  $y_t$  を

$$y_t = d_t \exp(y_t^*/2).$$

と復元することができる．この変数変換を行うことにより  $y_t^*$  は  $h_t$  に関して線形となり (45) 式と (42) 式はよく知られる線形な状態空間モデルとなる．すでに述べたように  $\epsilon_t^*$  の分布は  $K = 10$  個の混合正規分布により近似することができるが， $\epsilon_t^*$  を所与とした  $\eta$  の条件付分布は

$$\eta_t | d_t, \epsilon_t^* \sim N(d_t \rho \sigma \exp(\epsilon_t^*/2), \sigma^2(1 - \rho^2)), \quad (46)$$

となり  $d_t$  に依存し，かつ期待値が  $\epsilon_t^*$  の非線形関数になっている<sup>38</sup>．そこで， $d$  を所与として以下のように 2 変量混合正規分布による近似を考える．

$$g(\epsilon_t^*, \eta_t | d_t) = \sum_{j=1}^K p_j f_N(\epsilon_t^* | m_j, v_j^2) f_N[\eta_t | d_t \rho \sigma \exp(m_j/2) \{a_j + b_j(\epsilon_t^* - m_j)\}, \sigma^2(1 - \rho^2)]. \quad (47)$$

表 3 では， $\exp(\epsilon_t^*/2)$  を線形近似する際の平均二乗誤差を小さくするように求めた  $(a_j, b_j)$  の値が与えられている．この近似の精度は高く， $\rho = -0.9$  においても真の確率密度関数とのずれは非常に小さい．この  $d$  の条件付 2 変量混合正規分布を用いて線形ガウス状態空間モデルを構成しサンプリングの効率性を大幅に改善することができる．

このとき MCMC 法によるサンプリングは，以下のように  $s = (s_1, \dots, s_T)$  と  $\{\theta = (\phi, \sigma^2, \rho), \mu, \mathbf{h}\}$  の 2 つのブロックに分けて行う． $\mathbf{y}^* = \{y_t^*\}_{t=1}^T$ ， $\mathbf{d} = \{d_t\}_{t=1}^T$ ，事前確率密度関数を  $\pi(\theta), \pi(\mu)$  とおく．特に  $\mu$  の事前分布には  $\mu \sim \mathcal{N}(\mu_0, \sigma_{\mu_0}^2)$  を仮定する．

Step 1.  $s | \theta, \mu, \mathbf{h}, \mathbf{y}^*, \mathbf{d}$  を発生する

Step 2.  $(\theta, \mu, \mathbf{h}) | s, \mathbf{y}^*, \mathbf{d}$  を以下のように発生する．

(a)  $\theta | s, \mathbf{y}^*, \mathbf{d}$  を発生する．

<sup>38</sup>非対称性を考慮しない ( $\rho = 0$ ) ときには  $\eta_t$  の条件付分布は  $y_t$  の符号  $d_t$  に依存しない．また，このとき  $d_t$  は  $y_t^*$  とは独立になるので，パラメータの推定には  $y_t^*$  だけを考えればよい．

$j$	$p_j$	$m_j$	$v_j^2$	$a_j$	$b_j$
1	0.00609	1.92677	0.11265	1.01418	0.50710
2	0.04775	1.34744	0.17788	1.02248	0.51124
3	0.13057	0.73504	0.26768	1.03403	0.51701
4	0.20674	0.02266	0.40611	1.05207	0.52604
5	0.22715	-0.85173	0.62699	1.08153	0.54076
6	0.18842	-1.97278	0.98583	1.13114	0.56557
7	0.12047	-3.46788	1.57469	1.21754	0.60877
8	0.05591	-5.55246	2.54498	1.37454	0.68728
9	0.01575	-8.68384	4.16591	1.68327	0.84163
10	0.00115	-14.65000	7.33342	2.50097	1.25049

表 3:  $(p_j, m_j, v_j^2, a_j, b_j)$  の選択. Omori *et al.* (2006)

- (b)  $\mu|\theta, s, \mathbf{y}^*, d$  を発生する  
(c)  $h|\mu, \theta, s, \mathbf{y}^*, d$  を発生する

以下では, 各ステップについて説明していくこととする.

Step 1.  $s_t$  の事後分布は  $\theta, \mu, h, \mathbf{y}^*, d$  を所与として以下ようになる.

$$\begin{aligned} \pi(s_t = i | h, \mu, \theta, \mathbf{y}^*, d) \\ \propto \Pr(s_t = i) v_i^{-1} \exp \left\{ -\frac{(y_t^* - h_t - m_i)^2}{2v_i^2} \right. \\ \left. - \frac{[(h_{t+1} - \mu) - \phi(h_t - \mu) - d_t \rho \sigma \exp(m_i/2) \{a_i + b_i (y_t^* - h_t - m_i)\}]^2}{\sigma^2(1 - \rho^2)} \right\}. \end{aligned}$$

したがって, この離散分布からの乱数を発生させればよい.

Step 2a.  $s, \mathbf{y}^*, d$  を所与とした  $\theta$  の尤度関数を  $g(\mathbf{y}^* | d, s, \theta)$  とおくと事後密度関数は  $(\mu, h)$  に関して積分した周辺事後密度

$$\pi(\theta | s, \mathbf{y}^*, d) \propto g(\mathbf{y}^* | d, s, \theta) \pi(\theta),$$

から MH アルゴリズムを用いてサンプリングすればよい. 提案分布には事後確率密度  $\pi(\theta | s, \mathbf{y}^*, d)$  を近似する (切断) 正規分布を用いる.  $\hat{\theta} = (\hat{\phi}, \hat{\sigma}_\eta^2, \hat{\rho})'$  を  $\pi(\theta | s, \mathbf{y}^*, d)$  のモード (またはその近似値) とすると, 候補  $\gamma^*$  を領域  $R = \{\gamma : |\phi| < 1, \sigma^2 > 0, |\rho| < 1\}$  で切断された正規分布  $TN_R(\hat{\theta}, \Sigma_*)$  から発生させる. ただし

$$\Sigma_*^{-1} = - \left. \frac{\partial^2 \log g(\mathbf{y}^* | d, s, \theta) \pi(\theta)}{\partial \theta \partial \theta'} \right|_{\theta = \hat{\theta}},$$

あるいは, パラメータの制約をはずすため  $\theta_1 = \log(1 + \phi) - \log(1 - \phi), \theta_2 = \log \sigma_1^2, \theta_3 = \log(1 + \rho) - \log(1 - \rho)$  と変換してから, 提案分布として正規分布を構成してもよい. また, 上のヘッシアン行列が負値定符号ではない場合には (たとえば  $|\hat{\rho}| \approx 1$  のとき), 提案分布を  $TN_R(\hat{\theta}, c_0 I)$  ( $c_0$  は大きな値をとる定数) とすればよい.

この  $g(\mathbf{y}^* | d, s, \theta)$  を計算するには以下のように拡大カルマン・フィルタを用いればよい.

(1)  $a_1 = \mu, P_1 = \sigma^2/(1 - \phi^2)$  とする .

(2)  $a_1^* = 0, A_1^* = -1, q_1 = \sigma_{\mu 0}^{-2}\mu_0, Q_1 = \sigma_{\mu 0}^{-2}$  とする .

(3)  $t = 1, 2, \dots, T$  のとき

(a) カルマン・フィルタ

$$a_{t+1} = (1 - \phi)\mu + d_t \rho \sigma a_{s_t} \exp(m_{s_t}/2) + \phi a_t + K_t e_t, \quad (48)$$

$$P_{t+1} = \phi P_t L_t + H_t J_t', \quad (49)$$

を計算する . ただし

$$e_t = y_t^* - m_{s_t} - a_t, \quad D_t = P_t + v_{s_t}^2, \quad K_t = (\phi P_t + H_t G_t') D_t^{-1},$$

$$L_t = \phi - K_t, \quad J_t = H_t - K_t G_t,$$

$$G_t = (v_{s_t}, 0), \quad H_t = \left( d_t \rho \sigma b_{s_t} v_{s_t} \exp(m_{s_t}/2), \sigma \sqrt{1 - \rho^2} \right).$$

(b) 拡大 (augmented) カルマン・フィルタ

$$f_t = y_t^* - m_{s_t} - a_t^*, \quad a_{t+1}^* = d_t \rho \sigma a_{s_t} \exp(m_{s_t}/2) + \phi a_t^* + K_t f_t,$$

$$F_t = -A_t^*, \quad A_{t+1}^* = -(1 - \phi) + \phi A_t^* + K_t F_t,$$

を計算し ,

$$q_{t+1} = q_t + F_t f_t / D_t, \quad Q_{t+1} = Q_t + F_t^2 / D_t,$$

とする .

(4) 対数尤度  $\log g$  を

$$\begin{aligned} & \log g(\mathbf{y}^* | \mathbf{d}, \mathbf{s}, \boldsymbol{\theta}) \\ &= \text{定数} - \frac{1}{2} \left\{ \sum_{t=1}^T \log |D_t| + \log |Q_{T+1}| + \sum_{t=1}^T f_t^2 / D_t + \mu_0^2 \sigma_{\mu 0}^{-2} - q_{T+1}^2 / Q_{T+1} \right\}. \end{aligned}$$

とする .

Step 2b.  $\mathbf{d}, \mathbf{s}, \boldsymbol{\theta}, \mathbf{y}^*$  を所与としたときの  $\mu$  の事後分布は  $\mathcal{N}(Q_{T+1}^{-1} q_{T+1}, Q_{T+1}^{-1})$  となる (ただし  $q_{T+1}, Q_{T+1}$  は採択された標本  $\boldsymbol{\theta}$  を所与として Step 1a において求めたもの). したがって  $\mu \sim \mathcal{N}(Q_{T+1}^{-1} q_{T+1}, Q_{T+1}^{-1})$  を発生すればよい .

Step 2c.  $(\boldsymbol{\theta}, \mu, \mathbf{s}, \mathbf{y}^*, \mathbf{d})$  及び  $h$  を所与として  $h$  は次のようにシミュレーション・スムーザを用いてサンプリングすることができる .

(1) 以下のカルマン・フィルタを行い ,  $\{e_t, D_t, J_t, L_t\}_{t=1}^T$  を保存する .

(a)  $a_1 = \mu, P_1 = \sigma^2/(1 - \phi^2)$  とする .



(b)  $t = 1, 2, \dots, T - 1$  のとき

$$a_{t+1} = (1 - \phi)\mu + d_t \rho \sigma a_{s_t} \exp(m_{s_t}/2) + \phi a_t + K_t e_t, \quad (50)$$

$$P_{t+1} = \phi P_t L_t + H_t J_t', \quad (51)$$

を計算する．ただし

$$e_t = y_t^* - m_{s_t} - a_t, \quad D_t = P_t + v_{s_t}^2, \quad K_t = (\phi P_t + H_t G_t') D_t^{-1},$$

$$L_t = \phi - K_t, \quad J_t = H_t - K_t G_t,$$

$$G_t = (v_{s_t}, 0), \quad H_t = \left( d_t \rho \sigma b_{s_t} v_{s_t} \exp(m_{s_t}/2), \sigma \sqrt{1 - \rho^2} \right),$$

(2) 以下のシミュレーション・スムーザを行い,  $\{\xi_t\}_{t=0}^{T-1}$  を保存する.

(a)  $r_T = 0$  と  $U_T = 0$  として,  $t = T, T - 1, \dots, 1$  の順に

$$C_t = H_t(I - v_{s_t}^2 D_t^{-1} - J_t' U_t J_t) H_t', \quad \kappa_t \sim \mathcal{N}(0, C_t),$$

$$V_t = H_t(G_t' D_t^{-1} + J_t' U_t L_t),$$

$$r_{t-1} = D_t^{-1} e_t + L_t r_t - V_t C_t^{-1} \kappa_t,$$

$$U_{t-1} = D_t^{-1} + L_t^2 U_t + V_t^2 C_t^{-1},$$

を逐次的に計算して行き,  $\{\kappa_t\}_{t=1}^T, \{r_t\}_{t=0}^T$  を求める．

(b) さらに  $\kappa_0$  を

$$C_0 = \frac{\sigma^2}{1 - \phi^2} - \frac{\sigma^4}{(1 - \phi^2)^2} U_0, \quad \kappa_0 \sim \mathcal{N}(0, C_0),$$

と求める．

(c) (a)(b) で得られた  $\{\kappa_t\}_{t=0}^{T-1}, \{r_t\}_{t=0}^{T-1}$  より

$$\xi_t = H_t(G_t' D_t^{-1} e_t + J_t' r_t) + \kappa_t, \quad t = 1, \dots, T - 1,$$

$$\xi_0 = \frac{\sigma^2}{1 - \phi^2} r_0 + \kappa_0,$$

を求める．このとき  $\xi = (\xi_0, \xi_1, \dots, \xi_{T-1})$  は  $\eta = (\eta_0, \eta_1, \dots, \eta_{T-1})$  の事後分布からの確率標本となる．

(3)  $\xi$  を用いて  $h = (h_1, \dots, h_T)$  の事後分布からの確率標本を

$$h_{t+1} = (1 - \phi)\mu + d_t \rho \sigma a_{s_t} \exp(m_{s_t}/2) + \phi h_t + \xi_t, \quad t = 1, \dots, T - 1,$$

$$h_1 = \mu + \xi_0,$$

とすればよい．

### 5.1.2 mixture sampler における近似の修正

mixture sampler のサンプリングでは、

$$\xi_t^j = y_t^* - h_t^j, \quad \eta_t^j = (h_{t+1}^j - \mu^j) - \phi^j(h_t^j - \mu^j).$$

として真の二変量確率密度  $f(\xi_t, \eta_t | d_t, \theta)$  を混合正規確率密度  $g(\xi_t, \eta_t | d_t, \theta)$  により近似をしており、得られる標本

$$h^j, \mu^j, \theta^j, \quad j = 1, 2, \dots, M,$$

は近似事後確率密度  $g(h, \mu, \theta | y^*, d)$  からの標本となっている。近似のずれを修正して真の事後確率密度  $f(h, \mu, \theta | y^*, d)$  からの標本を得るためには

$$w_j^* = \prod_{t=1}^T \frac{f(\xi_t^j, \eta_t^j | d_t, \theta^j)}{g(\xi_t^j, \eta_t^j | d_t, \theta^j)}, \quad j = 1, 2, \dots, M,$$

を求めて、重み  $w_j$  を

$$w_j = \frac{w_j^*}{\sum_{i=1}^M w_i^*}.$$

とし  $(h^j, \mu^j, \theta^j)$  を  $w_j$  に比例して再び標本抽出すればよい。また、この重みを用いれば  $\theta$  の関数  $g(\theta)$  の事後平均は

$$\hat{E}[g(\theta)] = \sum_{j=1}^M w_j g(\theta^j).$$

として求めればよい。実証分析では  $\rho = -0.4 \sim -0.5$  程度であり、 $w_j$  の値はほぼ  $1/M$  でその分散も小さい。したがって後の実証分析の節で見るとように  $w_j$  を用いて重みづけをしても得られる結果にほとんど違いはない。

### 5.1.3 モンテカルロ・フィルタによる周辺尤度の計算

周辺尤度の推定は周辺尤度の恒等式を用いて行う。事前密度関数は既知の確率密度関数にパラメータの事後平均を代入して求めることができ、また事後確率密度関数は Chib and Jeliazkov (2001) によって計算すればよい。一方、尤度関数については以下のように補助粒子フィルタ (auxiliary particle filter) と呼ばれるモンテカルロ・フィルタを用いて計算すればよい。

まず観測方程式と状態方程式の確率密度関数を

$$f(y_t | h_t) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} h_t - \frac{1}{2} y_t^2 \exp(-h_t) \right\} \quad (52)$$

$$f(h_{t+1} | y_t, h_t, \mu, \theta) = \frac{1}{\sqrt{2\pi(1-\rho^2)\sigma}} \exp \left\{ -\frac{(h_{t+1} - \mu_{t+1})^2}{2(1-\rho^2)\sigma^2} \right\}, \quad (53)$$

(ただし  $\mu_{t+1} = \mu + \phi(h_t - \mu) + \rho\sigma \exp(-h_t/2)y_t$ )

とおく．次に粒子フィルタの計算で，候補となる粒子を発生させる重点関数として

$$\begin{aligned} g(h_{t+1}, h_t^i | \mathbf{y}_{t+1}, \mu, \boldsymbol{\theta}) &\propto f(y_{t+1} | \mu_{t+1}^i) f(h_{t+1} | y_t, h_t^i, \mu, \boldsymbol{\theta}) \hat{f}(h_t^i | \mathbf{y}_t, \mu, \boldsymbol{\theta}) \\ &\propto f(h_{t+1} | y_t, h_t^i, \mu, \boldsymbol{\theta}) g(h_t^i | \mathbf{y}_{t+1}, \mu, \boldsymbol{\theta}) \end{aligned} \quad (54)$$

をとる．ただし

$$\begin{aligned} g(h_t^i | \mathbf{y}_{t+1}, \mu, \boldsymbol{\theta}) &= \frac{f(y_{t+1} | \mu_{t+1}^i) \hat{f}(h_t^i | \mathbf{y}_t, \mu, \boldsymbol{\theta})}{\sum_{j=1}^I f(y_{t+1} | \mu_{t+1}^j) \hat{f}(h_t^j | \mathbf{y}_t, \mu, \boldsymbol{\theta})}, \\ f(y_{t+1} | \mu_{t+1}^i) &= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \mu_{t+1}^i - \frac{1}{2} y_{t+1}^2 \exp(-\mu_{t+1}^i) \right\}, \\ \mu_{t+1}^i &= \mu + \phi(h_t^i - \mu) + \rho \sigma \exp(-h_t^i/2) y_t, \\ \mathbf{y}_t &= (y_1, \dots, y_t), \end{aligned}$$

である．このとき以下の手順で補助粒子フィルタを行い，尤度を求めることができる．

(1)  $t = 1$  のとき  $h_1^i \sim \mathcal{N}(\mu, \sigma^2/(1 - \phi^2))$   $i = 1, 2, \dots, I$  を発生させる．

(a)  $w_i = f(y_1 | h_1^i)$  と  $W_i = F(y_1 | h_1^i)$ , (ただし  $F$  は  $h_t$  を所与としたときの  $y_t$  の分布関数) を計算し，

$$\bar{w}_1 = \frac{1}{I} \sum_{i=1}^I w_i, \quad \bar{W}_1 = \frac{1}{I} \sum_{i=1}^I W_i.$$

を保存する．

(b)  $\hat{f}(h_1^i | y_1, \mu, \boldsymbol{\theta}) = \pi_1^i = w_i / \sum_{j=1}^I w_j$ ,  $i = 1, 2, \dots, I$  とする．

(2) (54) 式の重点関数  $g(h_{t+1}, h_t | Y_{t+1}, \mu, \boldsymbol{\theta})$  を用いて  $(h_t^i, h_{t+1}^i)$  を発生させる ( $i = 1, \dots, I$ )

(a)

$$w_i = \frac{f(y_{t+1} | h_{t+1}^i) f(h_{t+1}^i | y_t, h_t^i, \mu, \boldsymbol{\theta}) \hat{f}(h_t^i | \mathbf{y}_t, \mu, \boldsymbol{\theta})}{g(h_{t+1}^i, h_t^i | \mathbf{y}_{t+1}, \mu, \boldsymbol{\theta})} = \frac{f(y_{t+1} | h_{t+1}^i) \hat{f}(h_t^i | \mathbf{y}_t, \mu, \boldsymbol{\theta})}{g(h_t^i | \mathbf{y}_{t+1}, \mu, \boldsymbol{\theta})},$$

$$W_i = \frac{F(y_{t+1} | h_{t+1}^i) \hat{f}(h_t^i | \mathbf{y}_t, \mu, \boldsymbol{\theta})}{g(h_t^i | \mathbf{y}_{t+1}, \mu, \boldsymbol{\theta})}, \quad i = 1, \dots, I,$$

を計算し，

$$\bar{w}_t = \frac{1}{I} \sum_{i=1}^I w_i, \quad \bar{W}_t = \frac{1}{I} \sum_{i=1}^I W_i. \quad (55)$$

を保存する．

(b)  $\hat{f}(h_{t+1}^i | \mathbf{y}_{t+1}, \mu, \boldsymbol{\theta}) = \pi_{t+1}^i = w_i / \sum_{j=1}^I w_j$ ,  $i = 1, 2, \dots, I$  とする．

(3)  $t$  を 1 増やして (2) に戻る．

$I \rightarrow \infty$  のとき,  $\bar{w}_{t+1}$ ,  $\bar{W}_{t+1}$  はそれぞれ予測確率密度関数, 予測分布関数に確率収束し

$$\bar{w}_{t+1} \xrightarrow{p} f(y_{t+1}|\mathbf{y}_t, \mu, \boldsymbol{\theta}), \quad \bar{W}_{t+1} \xrightarrow{p} F(y_{t+1}|\mathbf{y}_t, \mu, \boldsymbol{\theta})$$

となることが示せるので, 対数尤度関数の推定量  $\sum_{t=1}^T \log \bar{w}_t$  も

$$\sum_{t=1}^T \log \bar{w}_t \xrightarrow{p} \sum_{t=1}^T \log f(y_t|\mathbf{y}_{t-1}, \mu, \boldsymbol{\theta}),$$

である. またモデルが正しいときには  $2|\bar{W}_t - 1/2|$  が近似的に, 互いに独立に一様分布にしたがうので, これをモデルの特定化の診断に使うこともできる (例えば Kim, Shephard and Chib (1998), Shephard (1994b), Smith (1985), Rosenblatt (1952), Diebold, Gunther and Tay (1998) を参照).

## 5.2 誤差項の分布

資産価格変化率は正規分布よりも裾の厚い分布に従っている (尖度が3を越える) ことが古くから知られている (Mandelbrot 1963, Fama 1965). このことは,  $y_t$  が SV モデルに従っている場合, たとえ (10) 式の誤差項  $\epsilon_t$  が正規分布であっても,  $y_t$  の尖度は3を越える.<sup>39</sup> しかし, だからと言って,  $\epsilon_t$  の分布が標準正規分布で良いとは限らない.  $\epsilon_t$  に正規分布以外の分布も当てはめ, どの分布が最もフィットが良いかを分析することは重要である.

次の6節では,  $\epsilon_t$  の分布を学生t分布の  $t$  にした場合の推定も行っている<sup>40</sup>.

別の推定法を用いて SV モデルの誤差項  $\epsilon_t$  に正規分布以外の分布を当てはめているものに, Liesenfeld and Jung (2000) がある. そこでは, シミュレーションによる最尤法 (simulated maximum likelihood estimation; SML) を用いている. また, SV モデルではなく, ARCH 型モデルを使って同様の研究を行っているものに, Bollerslev, Engle and Nelson (1994), Watanabe (2001), 渡部 (2001, 2.4.2 節) がある. これらの研究では,  $\epsilon_t$  の分布に正規分布,  $t$  分布, 一般化誤差分布 (generalized error distribution; GED), およびそれらを含むより一般的な分布である一般化  $t$  分布を当てはめており,  $t$  分布が最も当てはまりが良いとの結果が得られている.

Chib, Nardari and Shephard (2002) では, 誤差項の分布を変えると同時に, ジャンプを加えたモデルを MCMC 法を用いてベイズ推定している. また, Jacquier, Polson and Rossi (2004) では, 誤差項の分布を変えると同時に, 以下で説明するボラティリティ変動の非対称性を考慮したモデルを MCMC 法を用いてベイズ推定している.

## 5.3 マルコフスイッチング SV モデル

So, Lam and Li (1998), Kalimipalli and Susmel (2004), Shibata and Watanabe (2004) らは, ボラティリティの平均  $\mu$  の値が一定ではなく, 高い時期 ( $S_t = 1$ ) と低い時期 ( $S_t = 0$ ) が

<sup>39</sup>詳しくは, 渡部 (2005) を参照のこと.

<sup>40</sup> $\epsilon_t$  の分布を学生t分布の  $t$  にした SV モデルの MCMC 法を用いたベイズ推定法については, 渡部 (2005) を参照のこと. また, その場合,  $t$  分布の自由度も事後分布からサンプリングしなければならないが,  $t$  分布の自由度の効率的なサンプリング法については, Watanabe (2001) を参照のこと.

あるものと考え、(2) 式を以下のように拡張している。

$$h_t = \mu_0 + \mu_1 S_t + \phi(h_{t-1} - \mu_0 - \mu_1 S_{t-1}) + \eta_t. \quad (56)$$

彼らは、Hamilton (1989) に従い、 $S_t$  はマルコフ過程に従うものと仮定している。

このモデルをマルコフ連鎖モンテカルロ法を用いてベイズ推定する場合には、 $S_T = (S_1, \dots, S_T)$  を  $f(S_T|h_T, \theta)$  からサンプリングしなければならないが、これは容易である。 $y_T$  が与えられると、(56) 式は Hamilton (1989) のマルコフスイッチングモデルになるので、Carter and Kohn (1994) や Chib (1996) の multi-move sampler により  $S_T = (S_1, \dots, S_T)$  を  $f(S_T|h_T, \theta)$  から 1 度にサンプリングできる。

上記の論文では、(56) 式を用いた場合、 $\phi$  の推定値が大幅に低下している。また、Kalimipalli and Susmel (2004) は予測パフォーマンスの比較により、Shibata and Watanabe (2004) は前節で説明した周辺尤度の比較やモデルの診断により、単純な SV モデルよりもマルコフスイッチング SV モデルの方が当てはまりが良いとの結果を得ている。

#### 5.4 取引高の導入

ボラティリティと取引高の間には正の相関があることが知られている (Karpoff 1987)。SV モデルを取引高を含める形で拡張したモデルに、Tauchen and Pitts (1983), Andersen (1996) による動学的 2 変量分布混合 (dynamic bivariate mixture; DBM) モデルがある。このモデルは、1 日のうちの 1 回 1 回の取引で生じる価格変化率と取引高は独立であるが、1 日の取引回数が増えることにより、日々のボラティリティと取引高との間に正の相関が生まれるとするモデルである。Tauchen and Pitts (1983) モデルでは、第  $t$  日の取引回数  $I_t$  が与えられると、価格変化率  $y_t$  と取引高  $v_t$  は次のような互いに独立な正規分布に従う。

$$y_t|I_t \sim \mathcal{N}(0, \sigma_r^2 I_t), \quad v_t|I_t \sim \mathcal{N}(\mu_v I_t, \sigma_v^2 I_t).$$

Andersen (1996) モデルでは、価格変化率の分布は同じで、取引高の分布が次のようなポアソン分布になる。

$$v_t|I_t \sim POI(m_0 + m_1 I_t).$$

いずれも、 $\log I_t$  が AR(1) モデルに従うものとし、 $v_t$  を無視すると、SV モデルになるので、これらのモデルは SV モデルを取引高を含める形で拡張したモデルになっている。そこで、このモデルのパラメーターもやはり最尤推定することが困難であり、また、このモデルは線形状態空間モデルに変換することができないので、mixture sampler は使えない。Watanabe (2000a, 2003) は multi-move sampler を用いてベイズ推定を行い、日経 225 先物の日次データでは、Tauchen and Pitts (1983) モデルは価格変化率の 2 乗の自己相関を過小評価し、Andersen (1996) モデルは取引高の自己相関を過小評価するとの結果を得ている。Lamoureux and Lastrapes (1994) は積率法 (method of moments; MM), Andersen (1996) は一般化積率法 (generalized method of moments; GMM), Liesenfeld (1998, 2001), Liesenfeld and Richard (2003) はシミュレーションによる最尤法 (simulated maximum likelihood estimation; SML) を用いて推定を行っている。

## 6 株式収益率のデータへの応用

この節では SV モデルを実際の株式収益率に応用した例を紹介する．データは TOPIX (東証株価指数, Tokyo Stock Price Index) の日次収益率  $y_t$  で、終値  $p_t$  の対数階差に 100 を乗じて  $y_t = 100 \times (\log p_t - \log p_{t-1})$  と計算したものである．使用した期間は 1998 年 1 月 5 日から 2002 年 12 月 30 日までの 1,232 営業日の日次収益率で、要約統計量は表 4 にまとめられている．これらのデータを使って非対称性のある SV モデルを mixture sampler によって推定し

TOPIX (1998/1/5 - 2002/12/30)

観測個数	平均	標準偏差	最大値	最小値	正值の個数	負値の個数
1,232	-0.0255	1.2839	5.3749	-5.6819	602	630

表 4: TOPIX 収益率の要約統計量 (Omori *et al.* (2006)).

た<sup>41</sup>．パラメータの事前分布は

$$\begin{aligned}\mu &\sim \mathcal{N}(0, 1), \quad \frac{\phi + 1}{2} \sim \text{Beta}(20, 1.5), \\ \sigma^2 &\sim \text{IG}\left(\frac{5}{2}, \frac{0.05}{2}\right), \quad \rho \sim U(-1, 1),\end{aligned}$$

とした．ただし  $\text{Beta}(a, b)$  は  $(a, b)$  を母数とするベータ分布,  $\text{IG}(a, b)$  は  $(a, b)$  を母数とする逆ガンマ分布,  $U(-1, 1)$  は区間  $(-1, 1)$  上の一様分布を表す． $\phi$  と  $\sigma^2$  の事前分布は Kim, Shephard and Chib (1998) と同様である．稼動検査期間 (burn-in period) として最初の 500 個を棄てて後、5,000 個のパラメータベクトルを発生させた．図 2 は標本自己相関関数、標本経路、事後確率密度関数である．標本自己相関関数はいずれのパラメータについても急速に減衰して定常分布への収束が早く、また標本経路も状態空間を万遍なく十分に訪れていることがわかる．

表 5 は推定された事後平均、事後標準偏差、95%信用区間、非効率性因子、相関行列である<sup>42</sup>．非効率性因子の値はどれも非常に小さく 10 前後であり、single move を用いた計算例では 100 ~ 2000 (Omori *et al.* (2006)) であるのとは対照的にサンプリングの高い効率性を示している． $\rho$  の事後平均は  $-0.36$  と負値であり、非対称性・レバレッジ効果の存在を示唆している． $\rho$  の 95%信用区間も  $[-0.59, -0.11]$  であり、 $\rho$  が負である事後確率は 0.95 よりも大きい．また事後分布の相関係数が  $\phi$  と  $\sigma$  で大きいので、これらのパラメータを同時にサンプリングする integration sampler が有効に機能していることがわかる．

表 6 は  $w_j$  を用いて重みづけすることにより真の事後分布の平均を推定し直した結果であるが、予想していたように結果はほとんど変わらない．実際、 $\log(w_j \times M)$  のヒストグラムを描いてみると図 3 のように、0 の周辺に集中しており、標準偏差も 0.34 と非常に小さい．

最後に非対称性・レバレッジ効果のあるモデルを含めたいくつかのモデルについて周辺尤度を計算して比較をう．推定したモデルは、(1) 分布の裾の厚いモデル (2) 重ね合わせモデル

<sup>41</sup>mixture sampler では、(24) 式の  $y_i^*$  を  $y_i^* = \log(y_i^2 + 0.0001)$  と計算している．

<sup>42</sup>非効率性因子を計算する際にはバンド幅を 100 として計算している．100 より大きいラグでは自己相関関数がほぼ 0 となるため、バンド幅をこれ以上大きくしても計算結果は変わらない．

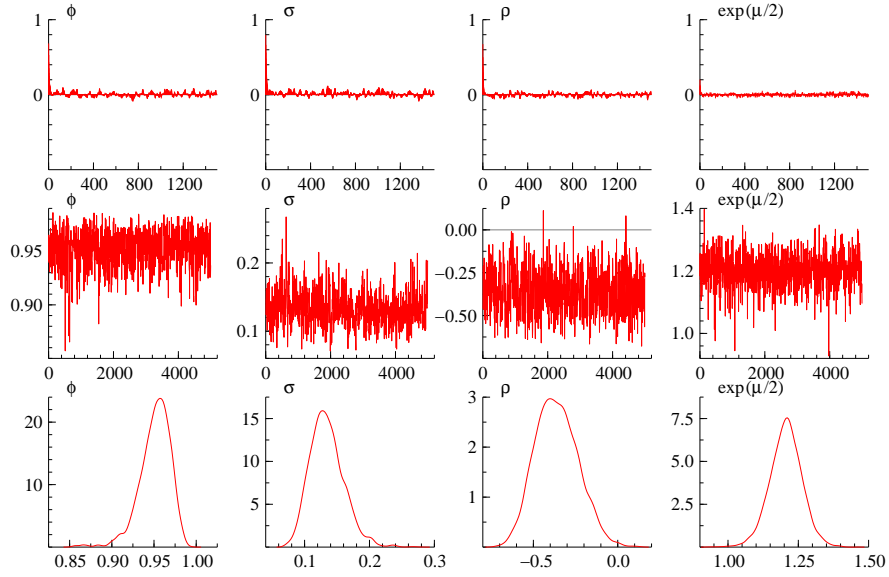


図 2: TOPIX の収益率を用いた推定結果. ( $\phi, \sigma, \rho, \beta = \exp(\mu/2)$ )  
 標本自己相関関数 (上段), 標本経路 (中段), 事後確率密度関数 (下段)

パラメータ	平均	標準偏差	95%信用区間	非効率性因子	相関係数			
$\phi$	0.9511	0.0185	[0.908, 0.980]	9.3	1	-.66	-.30	-.06
$\sigma$	0.1343	0.0262	[0.091, 0.193]	13.0		1	.19	-.08
$\rho$	-0.3617	0.1265	[-0.593, -0.107]	6.8			1	.13
$\beta$	1.2056	0.0573	[1.089, 1.318]	2.7				1

表 5: TOPIX データを用いた推定結果.

(superposition model) であり, (1) は現実のデータにおける分布の裾の厚さを反映したモデルであり, (2) は連続時間モデルに近いモデルとして知られている (例えば Shephard (1996), Engle and Lee (1999), Barndorff-Nielsen and Shephard (2001), Chernov *et al.* (2003)).

(1) 分布の裾の厚いモデル.

$$y_t = \exp(h_t/2) \sqrt{\lambda_t} \varepsilon_t, \quad \varepsilon_t \sim i.i.d. \mathcal{N}(0, 1), \quad (57)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t, \quad \eta_t \sim i.i.d. \mathcal{N}(0, \sigma^2). \quad (58)$$

ここで  $\lambda_t$  は互いに独立にガンマ分布  $\mathcal{G}(\nu/2, \nu/2)$  にしたがうとする. このとき  $\varepsilon_t = \sqrt{\lambda_t} \varepsilon_t$  は自由度  $\nu$  の  $t$  分布にしたがう裾の厚い分布となる.  $\nu$  の事前分布は  $\mathcal{G}(16, 0.8)$  とし, 平均 20, 分散 25 とした.

パラメータ	重みづけ後	
	平均	標準偏差
$\phi$	0.9512	0.0185
$\sigma$	0.1341	0.0264
$\rho$	-0.3578	0.1257
$\beta$	1.2052	0.0571

表 6: 重みづけ後に得られた事後分布の平均と標準偏差

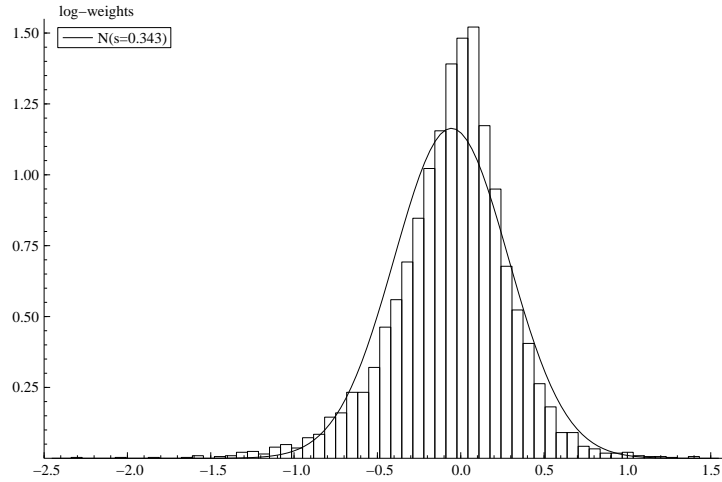


図 3: TOPIX データを用いた際の  $\log(w_k \times M)$  のヒストグラム。実線は同じ平均と分散をもつ正規分布の確率密度関数。

(2) 重ね合わせモデル (superposition model).

$$\begin{aligned} y_t &= \exp(h_t/2)\epsilon_t, \\ h_t &= \alpha_{1,t} + \alpha_{2,t}, \end{aligned}$$

ただし,  $\alpha_t = (\alpha_{1,t}, \alpha_{2,t})'$  は

$$\alpha_{t+1} = \begin{pmatrix} \mu \\ 0 \end{pmatrix} + \begin{pmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{pmatrix} \left( \alpha_t - \begin{pmatrix} \mu \\ 0 \end{pmatrix} \right) + \eta_t,$$

にしたがい,

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim i.i.d. \mathcal{N}(\mathbf{0}, \Sigma), \quad \Sigma = \begin{pmatrix} 1 & \rho_1\sigma_1 & \rho_2\sigma_2 \\ \rho_1\sigma_1 & \sigma_1^2 & 0 \\ \rho_2\sigma_2 & 0 & \sigma_2^2 \end{pmatrix}.$$

とする.  $\phi_1, \mu, \sigma_1^2, \rho_1$  の事前分布はこれまで同様に  $(\phi_1 + 1)/2 \sim \text{Beta}(20, 1.5), \mu \sim \mathcal{N}(0, 1), \sigma_1^2 \sim \text{IG}(5/2, 0.05/2), \rho_1 \sim U(-1, 1)$  とした.  $\phi_2$  については  $\phi_2 < \phi_1$  という制約のもとで切断される事前分布  $(\phi_2 + 1)/2 \sim \text{Beta}(10, 10)$  を,  $\rho_2$  については  $0 < \rho_1^2 + \rho_2^2 < 1$



という制約のもとで切断される  $(\rho_2 + 1)/2 \sim \text{Beta}(10, 10)$  をそれぞれ仮定した．また  $\sigma_2^2$  については  $\sigma_2^2 \sim \text{IG}(5/2, 0.05/2)$ . とした．

以上のモデルを組み合わせて以下の 5 つのモデルを推定し，周辺尤度による比較を行った．

Model 1 (SV) : 非対称性のない SV モデル.

Model 2 (SVt) : 非対称性のない SV モデルで， $\epsilon_t$  に  $t$  分布を仮定.

Model 3 (ASV) : 非対称性のある SV モデル.

Model 4 (ASVt) : 非対称性のある SV モデルで， $\epsilon_t$  に  $t$  分布を仮定.

Model 5 (SP) : 重ね合わせモデル.

表 7 は 5 つのモデルの周辺尤度，表 8 はベイズ・ファクターである．周辺尤度が最も高いのは Model 4 (ASVt) であるが，Model 3 (ASV) や Model 5 (SP) の 3 つのモデルはベイズ・ファクターからはほとんど違いは無く，ほぼ同等なモデルであるといえる．一方 Model 1 (SV) と Model 2 (SVt) の，Model 3, 4, 5 に対するベイズ・ファクターの値は  $-0.5$  より小さい．したがってこのデータでは非対称性を考慮したモデルのあてはまりが優れているが，その他の点では特に違いはないという結果となった．

	SV	SV-t	ASV	ASV-t	SP
周辺尤度	-2038.95	-2038.21	-2036.71	-2036.59	-2036.65
(標準誤差)	(0.16)	(0.28)	(0.10)	(0.14)	

表 7: 周辺尤度

	SV-t	ASV	ASV-t	SP
SV	-0.32	-0.97	-1.02	-1.00
SV-t		-0.65	-0.70	-0.68
ASV			-0.05	-0.03
ASV-t				0.03

表 8: ベイズ・ファクター.

(列のモデルに対する行のモデルのベイズ・ファクターの常用対数).

Nakajima and Omori (2006) では，観測方程式にジャンプの項を含めたモデルや GARCH モデル，EGARCH モデル，ジャンプに相関のあるモデルなどさまざまなモデルを 1970 年 1 月 1 日から 2003 年 12 月 31 日までの日次 S&P500 のデータと，1992 年 1 月 6 日から 2004 年 12 月 30 日までの日次の TOPIX のデータを用いて周辺尤度による比較を行っている．いずれの場合においても，非対称性のある SV モデルと（非対称性のある）重ね合わせ SV モデルがほぼ同様にあてはまりがよいということが報告されている．

## 7 今後の課題

本稿でサーベイしたように、SV モデルは、近年、推定法の開発が進み、それに伴って、モデルの拡張が行われるようになってきた。ところが、ARCH 型モデルと比べると、実際のデータへの応用はまだそれほど多いとは言えない。特に、実際のデータを用いて SV モデルやそれを発展させたモデルと ARCH 型モデルとを比較するという試みはこれまであまり行われておらず、今後の重要な課題と言えよう。

比較の方法としては、まず、SV モデル、ARCH 型モデル共に MCMC 法を用いてベイズ推定し、周辺尤度を比較するという方法が考えられうる<sup>43</sup>。ARCH 型モデルの MCMC 法を用いたベイズ推定法も、Bauwens and Lubrano (1998), Nakatsuma (2000), 三井・渡部 (2003) らによって提案されている。また、ARCH 型モデルは尤度が解析的に評価できるので、SV モデルより簡単に周辺尤度を計算できる。

金融実務の観点からは、ボラティリティの推定値や予測値、オプション価格、Value at Risk (VaR) などを用いた比較も重要である<sup>44</sup>。Deb (1997) は、SV モデルからシミュレーションによって人工的に発生させたデータを用いてボラティリティを推定するというモンテカルロ実験を行い、SV モデルを一般化積率法 (GMM) や疑似最尤法 (QML) といった正しい尤度に基づかない簡単な方法で推定するよりも、ARCH 型モデルを最尤推定することによりボラティリティを推定した方がパフォーマンスが良いとの結果を得ており、興味深い。

最後に、近年、精度の高いボラティリティの推定量として注目を集めているものに Realized Volatility (RV) がある。これは、日中の例えば 5 分ごとのリターンの 2 乗を足し合わせたものである。RV は長期記憶過程に従っていることが知られており、その変動を表すモデルとしては ARFIMA (autoregressive fractionally integrated moving average) モデルがよく用いられる<sup>45</sup>。こうした RV のモデルと ARCH 型モデルや SV モデルとの比較も重要であり、現在、盛んに行われている (Giot and Laurent (2004), Koopman, Jungbacker and Hol (2005), 渡部・佐々木 (2006))。RV はボラティリティの推定値や予測値の精度を比較する場合のボラティリティの真の値の代理変数としても用いられる。古くは、 $y_t^2$  が代理変数として用いられてきたが (渡部 (2000 2.3.3 節) 参照)、(10) 式からわかるように、 $y_t^2$  の変動は  $\exp(h_t)$  の変動だけでなく  $\epsilon_t^2$  の変動も含んでいるため、Andersen and Bollerslev (1998) は、 $y_t^2$  を代理変数として用いると ARCH 型モデルや SV モデルのボラティリティの予測パフォーマンスを過小評価してしまうことを示し、 $y_t^2$  ではなく、より精度の高いボラティリティの推定値である RV を使うことを提案している。また、Hansen and Lunde (2006) は代理変数を  $y_t^2$  にすると、予測精度の低いモデルを高いモデルとして間違えて選択してしまう可能性があることを示しており、やはり RV を使うことを提案している。

<sup>43</sup> そうした比較を行っているものに、Kim, Shephard and Chib (1998), Nakajima and Omori (2006) がある。

<sup>44</sup> 三井・渡部 (2003) は、GARCH モデルを MCMC 法を使ってベイズ推定し、オプション価格を計算している。VaR については、Giot and Laurent (2004), 渡部・佐々木 (2006) を参照のこと。

<sup>45</sup> 長期記憶過程や ARFIMA モデルについては、矢島 (2003), 渡部・佐々木 (2006) を参照のこと。RV のその他のモデルには Corsi (2004) の HAR (heterogeneous interval autoregressive) モデルや Barndorff-Nielsen and Shephard (2001, 2002) の UC (unobserved component) モデルなどがある。

## 参考文献

- 伊庭幸人・種村正美・大森裕浩・和合肇・佐藤整尚・高橋明彦 (2005), 『計算統計 II-マルコフ連鎖モンテカルロ法とその周辺-』東洋経済.
- 大森裕浩 (2001), 「マルコフ連鎖モンテカルロ法の最近の展開」『日本統計学会誌』 **31**, 305-344.
- 小西貞則・北川 源四郎 (2004) 『情報量規準』朝倉書店.
- 中妻照雄 (2003), 『ファイナンスのための MCMC 法によるベイズ分析』三菱経済研究所.
- 三井秀俊・渡部敏明 (2003), 「ベイズ推定法による GARCH オプション価格付けモデルの分析」『日本統計学会誌』 **33**, 307-324.
- 矢島美寛 (2003), 「長期記憶モデルをもつ時系列モデル」刈屋武昭・矢島美寛・田中勝人・竹内啓著『経済時系列の統計 その理論的基礎』第 II 部, 岩波書店, 103-202.
- 和合肇編著 (2005) 『ベイズ計量経済分析 マルコフ連鎖モンテカルロ法とその応用』東洋経済新報社.
- 渡部敏明 (2000), 『ボラティリティ変動モデル』朝倉書店.
- 渡部敏明 (2005), 「マルチ・ムーブ・サンプラーを用いた確率的ボラティリティ変動モデルのベイズ推定法」和合肇編著『ベイズ計量経済分析 マルコフ連鎖モンテカルロ法とその応用』第 9 章, PP.259-294, 東洋経済新報社.
- 渡部敏明・佐々木浩二 (2006), 「ARCH 型モデルと “Realized Volatility” によるボラティリティ予測とバリュエーション・アット・リスク」『金融研究』第 25 巻別冊第 2 号, 39-74.
- Andersen, T. G. (1996), “Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility,” *Journal of Finance*, **51**, 169-204.
- Andersen, T. G., and T. Bollerslev (1998), “Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts,” *International Economic Review*, **39**, 885-905.
- Ando, T. (2006), “Bayesian Predictive Information Criterion for the Evaluation of Hierarchical Bayesian and Empirical Bayes Models,” *Biometrika. forthcoming*.
- Barndorff-Nielsen, O. E. and N. Shephard (2001), “Non-Gaussian Ornstein-Uhlenbeck-based Models and Some of Their Uses in Financial Economics (with discussion),” *Journal of the Royal Statistical Society, Ser. B*, **63**, 167-241.
- Barndorff-Nielsen, O. E. and N. Shephard (2002), “Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility Models,” *Journal of the Royal Statistical Society, Ser. B*, **64**, 253-280.
- Bauwens, L., and M. Lubrano (1998), “Bayesian Inference on GARCH Models using the Gibbs Sampler,” *Econometrics Journal*, **1**, c23-c46.
- Bauwens, L., and M. Lubrano (2002), “Bayesian Option Pricing Using Asymmetric GARCH,” *Journal of Empirical Finance*, **9**, 312-342.
- Berger, J. O. and J. Mortera (1999), “Default Bayes Factors of Nonnested Hypothesis Testing,” *Journal of the American Statistical Association*, **94**, 542-554.
- Berger, J. O. and L. R. Pericchi (1996), “The Intrinsic Bayes Factor for Linear Models,” in *Bayesian Statistics*, **5**, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp.25-44. Oxford: Oxford University Press.
- Black, F. (1976), “Studies of Stock Market Volatility Changes,” *1976 Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 177-181.
- Bollerslev, T. (1986), “Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, **31**, 307-327.

- Bollerslev, T., R. F. Engle, and D.B. Nelson (1994), "ARCH Models," in *The Handbook of Econometrics 4*, eds. R. F. Engle and D. McFadden, pp.2959–3038, Amsterdam: North-Holland.
- Brooks, S. P. and G. O. Roberts (1999), "On Quantile Estimation and Markov Chain Monte Carlo Convergence," *Biometrika*, **86**, 710-717.
- Chen, M.-H., Q.-M. Shao and J. G. Ibrahim (2000), *Monte Carlo Methods in Bayesian Computation*. Springer: New York.
- Chib, S., and E. Greenberg (1994), "Bayes Inference for Regression Models with ARMA(p,q) Errors," *Journal of Econometrics*, **64**, 183–206.
- Carter, C. K., and R. Kohn (1994), "On Gibbs Sampling for State Space Models," *Biometrika*, **81**, 541–553.
- Chernov, M., A. R. Gallant, E. Ghysels and G. Tauchen (2003), "Alternative Models of Stock Price Dynamics," *Journal of Econometrics*, **116**, 225-257.
- Chib, S. (1995), "Marginal Likelihood from the Gibbs Output," *Journal of the American Statistical Association*, **90**, 1313–1321.
- Chib, S. (1996), "Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models," *Journal of Econometrics*, **75**, 79–97.
- Chib, S. (2001), "Markov Chain Monte Carlo Methods: Computation and Inference" in *Handbook of Econometrics*, **5**, eds J.J. Heckman and E. Leamer, North Holland, Amsterdam, 3569-3649.
- Chib, S., and Greenberg, E. (1994), "Bayes Inference for Regression Models with ARMA(p,q) Errors," *Journal of Econometrics*, **64**, 183–206.
- Chib, S., and E. Greenberg (1995), "Understanding the Metropolis-Hastings Algorithm," *American Statistician*, **49**, 327–335.
- Chib, S., F. Nardari and N. Shephard (2002), "Markov Chain Monte Carlo Methods for Generalized Stochastic Volatility Models," *Journal of Econometrics*, **108**, 281–316.
- Chib, S., and I. Jeliazkov (2001), "Marginal Likelihood from the Metropolis-Hastings Output," *Journal of the American Statistical Association*, **96**, 270–281.
- Chib, S., and I. Jeliazkov (2005), "Accept-Reject Metropolis-Hastings Sampling and Marginal Likelihood Estimation," *Statistica Neerlandica*, **59**, 30-44.
- Christie, A. A. (1982), "The Stochastic Behavior of Common Stock Variances: Value, Leverage, and Interest Rate Effects," *Journal of Financial Economics*, **10**, 407-432.
- Corsi, F. (2004), "A Simple Long Memory Model of Realized Volatility," Working Paper, Institute of Finance, University of Lugano.
- Cowles, M. K. and B. P. Carlin (1996), "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review," *Journal of the American Statistical Association*, **91**, 883-904.
- Cowles, M. K., G. O. Roberts and J. S. Rosenthal (1999), "Possible Bias Induced by MCMC Convergence Diagnostics," *Journal of Statistical Computation and Simulation*, **64**, 87-104.
- Danielsson, J. (1994), "Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood," *Journal of Econometrics*, **64**, 375–400.
- Danielsson, J., and J.-F. Richard (1993), "Accerlated Gaussian Importance Sampler with Application to Dynamic Latent Variable Models," *Journal of Applied Econometrics*, **8**, 153–174.
- de Jong, P. (1991), "The Diffuse Kalman Filter," *Annals of Statistics*, **19**, 1073-1083.

- de Jong, P. and N. Shephard (1995), "The Simulation Smoother for Time Series Models," *Biometrika*, **82**, 339-350.
- Deb, P. (1997), "Finite Sample Properties of the ARCH Class of Models with Stochastic Volatility," *Economics Letters*, **55**, 27-34.
- Diebold, F. X., T. A. Gunther and T. S. Tay (1998), "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, **39**, 863-883.
- Durbin, J., and S. J. Koopman (2001), *Time Series Analysis by State Space Models*, Oxford: Oxford University Press.
- Durbin, J., and S. J. Koopman (2002), "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis," *Biometrika*, **89**, 603-616.
- Engle, R. F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, **50**, 987-1007.
- Engle, R. F. and G. G. J. Lee (1999), "A Permanent and Transitory Component Model of Stock Return Volatility," in Engle, R. F. and H. White (Eds.), *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, 475-497. Oxford: Oxford University Press.
- Fama, E. F. (1965), "The Behavior of Stock Market Prices," *Journal of Business*, **38**, pp.34-105.
- Frühwirth-Schnatter, S. (2004), "Estimating Marginal Likelihoods for Mixture and Markov Switching Models using Bridge Sampling Techniques," *The Econometrics Journal*, **7**, 143-167.
- Frühwirth-Schnatter, S. (2006), *Finite Mixture and Markov Switching Models*. Springer.
- Gelfand, A. E. and A. F. M. Smith (1990), "Sampling-based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, **85**, 398-409.
- Gelman, A. (1996), "Inference and Monitoring Convergence," in *Markov Chain Monte Carlo in Practice*, Ed. W.R. Gilks, S. Richardson and D. J. Spiegelhalter, pp131-143. Chapman & Hall: London.
- Gelman, A., W. R. Gilks and G. O. Roberts (1996), "Efficient Metropolis Jumping Rules", in *Bayesian Statistics*, **5**, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp.599-608. Oxford: Oxford University Press.
- Gelman, A. and X.-L. Meng (1998), "Simulating Normalizing Constants: From Importance Sampling to Bridge Sampling to Path Sampling," *Statistical Science*, **13**, 163-185.
- Gelman, A. and D. B. Rubin (1992). "Inference from Iterative Simulation Using Multiple Sequences (with discussion)," *Statistical Science*, **7**, 457-511.
- Geman, S. and D. Geman (1984), " Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images, *Institute of Electrical and Electronics Engineers, Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.
- Geweke, J. (1992), "Evaluating the Accuracy of Sampling-based Approaches to the Calculation of Posterior Moments," in *Bayesian Statistics*, **4**, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp.169-193. Oxford: Oxford University Press.
- Geweke, J. (2004), "Getting It Right: Joint Distribution Tests of Posterior Simulators," *Journal of the American Statistical Association*, **99**, 799-804.
- Ghysels, E., A. C. Harvey and E. Renault (1996), "Stochastic volatility," in *Statistical Methods in Finance*, eds. C.R. Rao and G.S. Maddala, pp.119-191, Amsterdam: North-Holland.

- Giot, P. and S. Laurent (2004), "Modelling Daily Value-at-Risk Using Realized Volatility and ARCH Type Models," *Journal of Empirical Finance*, **11**, 379–398.
- Gordon, S., L. Samson and B. Carmichael (1996), "Bayesian Estimation of Stochastic Discount Factors," *Journal of Business & Economic Statistics*, **14**, 412–420.
- Hamilton, J. D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, **57**, 357–384.
- Han, C. and P.C. Carlin (2001), "Markov Chain Monte Carlo Methods for Computing Bayes Factors: A Comparative Review," *Journal of the American Statistical Association*, **96**, pp. 1122–1132.
- Harvey, A. C. (1989), *Forecasting Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.
- Harvey, A. C., E. Ruiz and N. Shephard (1994), "Multivariate Stochastic Variance Models," *Review of Economic Studies*, **61**, 247–264.
- Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications", *Biometrika*, **57**, 97–109.
- Hansen, P. R. and Lunde, A. (2006), "Consistent Ranking of Volatility Models," *Journal of Econometrics*, **131**, 97–121.
- Jacquier, E., N. Polson, and P. Rossi (1994), "Bayesian Analysis of Stochastic Volatility Models (with Discussion)," *Journal of Business & Economic Statistics*, **12**, 371–417.
- Jacquier, E., N. Polson, and P. Rossi (2004), "Bayesian Analysis of Stochastic Volatility Models with Fat-tails and Correlated Errors," *Journal of Econometrics*, **122**, 185–212.
- Jeffreys, H. (1961), *Theory of Probability*, 3rd edition, Oxford University Press, New York.
- Kass, R. E. and A. E. Raftery (1995), "Bayes Factors," *Journal of the American Statistical Association*, **90**, 773–795.
- Kalimipalli, M., and R. Susumel (2004), "Regime-switching Stochastic Volatility and Short-term Interest Rates," *Journal of Empirical Finance*, **11**, 309–329.
- Kim, S., N. Shephard, and S. Chib (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH models", *Review of Economic Studies*, **65**, 361–393.
- Koopman, S. J., B. Jungbacker and E. Hol (2005), "Forecasting Daily Variability of the S&P 100 Stock Index using Historical, Realized and Implied Volatility Measurements," *Journal of Empirical Finance*, 445–475. **12**
- Lamoureux, C. G., and W. D. Lastrapes (1994), "Endogenous Trading Volume and Momentum in Stock-Return Volatility," *Journal of Business & Economic Statistics*, **12**, 253–260.
- Liesenfeld, R. (1998), "Dynamic Bivariate Mixture Models: Modeling the Behavior of Prices and Trading Volume," *Journal of Business & Economic Statistics*, **16**, 101–109.
- Liesenfeld, R. (2001), "A Generalized Bivariate Mixture Model for Stock Price Volatility and Trading Volume," *Journal of Econometrics*, **104**, 141–178.
- Liesenfeld, R., and J.-F. Richard (2003), "Estimation of Dynamic Bivariate Mixture Models: Comments on Watanabe (2000)," *Journal of Business & Economic Statistics*, **21**, 570–576.
- Mahieu, R.J. and P.C. Schotman (1998), "An Empirical Application of Stochastic Volatility Models," *Journal of Applied Econometrics*, **13**, 333–360.
- Mandelbrot, B. (1963), "The Variation of Certain Speculative Prices," *Journal of Business*, **36**, pp.394–419.

- Melino, A., and S.M. Turnbull (1990), "Pricing Foreign Currency Options with Stochastic Volatility," *Journal of Econometrics*, **45**, 239–265.
- Meng, X.-L. and W. H. Wong (1996), "Simulating Ratios of Normalizing Constants via Simple Identity: A Theoretical Expolration," *Statistica Sinica*, **6**, 831-860.
- Mengersen, K. L., C. P. Robert and C. Guhenneuc-Jouyaux (1999), "MCMC Convergence Diagnostics:A Review," J.M. Bernardo, J.M. Berger, A.P.Dawid, and A.F.M. Smith (eds), *Bayesian Statistics*, **6**, 415–440,, Oxford University Press, Oxford.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), "Equations of State Calculations by Fast Computing Machines ", *Journal of Chemical Physics*, **21**, 1087-1091.
- Murdoch, D. J. and P. J. Green (1998), "Exact Sampling from a Continuous State Space," *Scandinavian Journal of Statistics*, **25**, 451-482.
- Nakajima, J. and Y. Omori (2006), "Bayesian Analysis for Jumps, Leverage and Heavy-tails in Stochastic Volatility, GARCH and EGARCH models," unpublished manuscript.
- Nelson, D. B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, **62**, 1-41.
- Newton, M. A. and A. E. Raftery (1994), "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap," *Journal of Royal Statistical Society, Ser. B*, **56**, 3-26.
- Ogata, Y. (1989), "A Monte-Carlo Method for High Dimensional Integration," *Numerische Mathematik*, **55**, 137-157.
- Ogata, Y. (1990), "A Monte-Carlo Method for an Objective Bayesian Procedure," *Annals of the Institute of Statistical Mathematics.*, **42**, 403-433.
- Omori, Y., S. Chib, N. Shephard and J. Nakajima (2006), "Stochastic Volatility with Leverage," *Journal of Econometrics. forthcoming*.
- Omori, Y., and T. Watanabe (2003), "Block Sampler and Posterior Mode Estimation for a Nonlinear and Non-Gaussian State-Space Model with Correlated Errors," CIRJE Discussion Paper F-221, Faculty of Economics, University of Tokyo.
- Pitt, M. K., and N. Shephard (1999), "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, **94**, 590–599.
- Propp, J. G. and B. M. Wilson (1996), "Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics," *Random Structures and Algorithms*, **9**, 223-252.
- Raftery, A. E. (1996), "Hypothesis Testing and Model Selection", in *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Spiegelhalter, eds, Chapman & Hall: London, pp.165-187.
- Raftery, A. E. and S. Lewis (1992), "How Many Iterations in the Gibbs Sampler?," J. M. Bernardo, J. M. Berger, A. P. Dawid, and A. F. M. Smith (eds), *Bayesian Statistics*, **4**, 763–773, Oxford University Press, Oxford.
- Raftery, A. E. and S. Lewis (1996), "Implementing MCMC", in *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Spiegelhalter, eds, Chapman & Hall: London, pp.165-187.
- Ripley, B. D. (1987), *Stochastic Simulation*, New York: John Wiley & Sons.
- Robert, C.P. and G. Casella (2004), *Monte Carlo Statistical Method*, 2nd ed., Springer, New York.

- Rosenblatt, M. (1952), "Remarks on a Multivariate Transformation," *Annals of Mathematical Statistics*, **23**, 470-472.
- Ruiz, E. (1994), "Quasi-maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Econometrics*, **63**, 289-306.
- Sandmann, G., and S.J. Koopman (1998), "Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood," *Journal of Econometrics*, **87**, 271-301.
- Shephard, N. (1994), "Partial Non-Gaussian State Space," *Biometrika*, **81**, 115-131.
- Shephard, N. (1996), "Statistical Aspects of ARCH and Stochastic Volatility," in Cox, D. R., Hinkley, D.V. and O. E. Barndorff-Nielsen (Eds.), *Time Series Models in Econometrics, Finance and Other Fields*, 1-67. London: Chapman & Hall.
- Shephard, N. (2004), *Stochastic Volatility: Selected Readings*, Oxford: Oxford University Press.
- Shephard, N., and M. K. Pitt (1997), "Likelihood Analysis of Non-Gaussian Measurement Time Series," *Biometrika*, **84**, 653-667.
- Shibata, M., and T. Watanabe (2004), "Bayesian Analysis of Markov Switching Stochastic Volatility Models," mimeo.
- Smith, J. Q. (1985), "Diagnostic Checks of Non-standard Time Series Models," *Journal of Forecasting*, **4**, 283-291.
- So, M. K. P., K. Lam and W. K. Li (1998), "A Stochastic Volatility Model with Markov Switching," *Journal of Business & Economic Statistics*, **16**, 244-253.
- So, M. K. P., and W. K. Li (1999), "Bayesian Unit-Root Testing in Stochastic Volatility Models," *Journal of Business & Economic Statistics*, **17**, 491-496.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin and A. van der Linde (2002), "Bayesian Measures of Model Complexity and Fit," *Journal of Royal Statistical Society, Ser. B*, **64**, 583-639.
- Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions (with Discussion)," *Annals of Statistics*, **22**, 1701-1762.
- Watanabe, T. (2000a), "Bayesian Analysis of Dynamic Bivariate Mixture Models: Can They Explain the Behavior of Returns and Trading Volume ?" *Journal of Business & Economic Statistics*, **18**, 199-210.
- Watanabe, T. (2000b), "Excess Kurtosis of Conditional Distribution for Daily Stock Returns: The Case of Japan," *Applied Economics Letters*, **7**, 353-355.
- Watanabe, T. (2001), "On Sampling the Degree-of-Freedom of Student-*t* Disturbances," *Statistics and Probability Letters*, **52**, 177-181.
- Watanabe, T. (2003), "The Estimation of Dynamic Bivariate Mixture Models: Reply to Liesenfeld and Richard Comments, " *Journal of Business & Economic Statistics*, **21**, 577-580.
- Watanabe, T., and Y. Omori (2004a), "A Multi-move Sampler for Estimating Non-Gaussian Time Series Models: Comments on Shephard and Pitt (1997)," COE Discussion Paper Series No.8, Faculty of Economics, Tokyo Metropolitan University.
- Watanabe, T., and Y. Omori (2004b), "A Multi-move Sampler for Estimating Non-Gaussian Time Series Models: Comments on Shephard & Pitt (1997)," *Biometrika*, **91**, 246-248.
- Wright, J.H. [1999] "Testing for a Unit Root in the Volatility of Asset Returns," *Journal of Applied Econometrics*, **14**, 309-318.
- Wu, G. (2001), "The Determinants of Asymmetric Volatility," *Review of Financial Studies*, **14**, 837-859.



Yu, J. (2005), "On Leverage in a Stochastic Volatility Model," *Journal of Econometrics*, **127**, 165-178.