A Review on Bayesian Financial Time Series Analysis

지도교수 이 상 열

이 논문을 이학석사 학위논문으로 제출함

2014년 2월

서울대학교 대학원

통계학과

이 희 석

이희석의 이학석사 학위논문을 인준함

2014년 2월

위 원 장 ____________________ (인)

부 위 원 장 ____________________ (인)

위 원 ____________________ (인)

이 학 석 사 학 위 논 문

A Review on Bayesian Finincial Time Series Analysis

금융 시계열 자료의 베이지안 분석에 대한 고찰

2014년 2월

서울대학교 대학원
통계학과
이 희 석

A Review on Bayesian Financial Time Series Analysis

by

Hee-seok Lee

A Thesis
submitted in fulfillment of the requirement
for the degree of
Master of Science
in Statistics

The Department of Statistics
College of Natural Sciences
Seoul National University
February, 2014

Abstract

This thesis introduces concepts and applications of Bayesian inference for financial time series. Bayesian inference is a method of statistical inference using Bayes' theorem to update prior beliefs as additional informations are observed. This allows us to use our prior beliefs of parameters and the Markov chain Monte Carlo method(MCMC) makes the analysis is relatively fast and simple.

In this thesis we introduce the time series Bayesian inference and the MCMC method, illustrate an example of estimating unknown parameters in threshold autoregressive(TAR) models with stochastic volatility(SV). Moreover, we apply TAR with SV model to a real data set and conduct a hypothesis test for model selection via using Bayes factor.

Key words : Bayesian time series, Mixed Prior, Markov chain Monte Carlo, Threshold autoregressive, Stochastic volatility, Bayes factor.

Student number : 2012-20230

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Advances of computational methodology have increased ability to solve complicated problem. They also extend the applicability of many existing statistical methods. Especially in the area of econometrics with many complex models, advantages of computational method have great importance.

Therefore, in this thesis we introduce one of the outstanding developments in computational methodology, Markov chain Monte Carlo(MCMC) method that are widely applicable in financial time series. First, in Chapter 2 we discuss Bayesian inference including general procedure and model comparison method via using Bayes factor. Bayesian inference provides us to insert prior beliefs of parameters before observing data. Second, in Chapter 3 we discuss algorithms of MCMC, which gives us the simple and fast way to calculate posterior density in Bayesian inference. Then, in Chapter 4 we focus on the financial time series problem. In particular, we carry out simulation and empirical studies based on former discussions.

For application, first we demonstrate simulation study of threshold autoregressive model(TAR) with stochastic volatility(SV). TAR model is one of tools to capture the nonlinearity of the financial time series. As well as mean structure, SV model is considered because modeling volatility has many applications in financial time series such as option trading or risk management. Also these two models are combined so this could be a good example of advantages of Bayesian inference via MCMC method in complicated financial time series problem. Second, we apply TAR with SV to a real data set and carry out a hypothesis testing via using Bayes factor for model selection problem.

Chapter 2

Bayesian Inference

Bayesian inference is a method of statistical inference by using Bayes' theorem to update prior beliefs when additional informations are observed. Bayesian inference is psychologically appealing because it allows us to insert our prior beliefs about parameters before data are observed.

Although it has a weakness of having subjective notion of probability, Bayesian approach has wide usage in many fields of statistics. In most cases solutions of two approaches are similar, even some cases Bayesian solutions might be advantageous.

2.1 General procedure

Bayesian inference is generally carried out in the following steps.

1. Choose the probability of parameter – prior distribution – before observing

data. Prior distribution reflects beliefs about parameter θ.

2. Choose model $f(x|\theta)$ reflects beliefs about x given θ.

3. After seeing data $X_1, \cdots, X_n$, use Bayes' theorem to calculate the posterior distribution $f(\theta|X_1, \cdots, X_n)$.

For the time series analysis, steps are similar.

Suppose we observe time series data $\boldsymbol{y} = (y_1, \cdots, y_n)$ from $\{y_t; t \geq 0\}$, collection of random variables over time. If we believe that y_t has a some density function $p(\cdot|\theta)$, our observation can be written as $p(\boldsymbol{y}|\theta)$. When we see this as the function of θ, we call it the likelihood function. Unlike Frequentist approaches that are mostly based on this likelihood function, Bayesian introduce pre-assumed beliefs called 'prior', $\pi(\theta)$

By Bayes' theorem,

$$\begin{aligned} p(\theta|y) &= \frac{p(y|\theta)\pi(\theta)}{\int_\Theta p(y|\theta)\pi(\theta)d\theta} \\ p(\theta|y) &\propto p(y|\theta)\pi(\theta) \end{aligned}$$

$p(\cdot|\theta)$ and $\pi(\cdot)$ denote likelihood and prior density function, respectively. Also we call $p(\theta|y)$ as posterior density. Bayesian inferences are based on this posterior distribution.

From the choice of prior distribution, it is classified into conjugate and nonconjugate analysis.

Conjugate Bayesian analysis

For convenience, the prior distribution often assumed to be conjugate, which is from same distribution family with corresponding posterior distribution. Advantages are its reasonable features and the simple calculation that may result in closed analytical form.

Nonconjugate Bayesian analysis

However, in many situations, there is no closed analytical form of posterior distribution against our desire.

It is hard to calculate integration in the denominator of posterior density function. It may need numerical approximation, or other methods.

Since the Markov chain Monte Carlo(MCMC) method introduced, it become possible to inference when we have no closed form of posterior distribution by obtaining sample draws from it. Next chapter, we see the algorithms of the MCMC method.

2.2 Bayes factor

For the model comparison or hypothesis testing, Bayesian approach uses Bayes factor(BF), the Bayesian version of the classical likelihood ratio test(LRT).

Consider two models or hypotheses H_1 and H_2 for given data y, Bayes factor is defined as,

$$BF = \frac{p(y|H_1)}{p(y|H_2)} = \frac{\int p(\theta_1|H_1)p(y|\theta_1, H_1)d\theta_1}{\int p(\theta_2|H_2)p(y|\theta_2, H_2)d\theta_2} \tag{2.1}$$

where θ_i s stand for parameters in H_i s.

Note Bayes' theorem says $p(H_j|y) \propto p(y|H_j)p(H_j)$. Therefore we can see that *posterior odds ratio* $\propto$ *BF* $\times$ *prior odds ratio.* The BF can be translated as ratio of the posterior odds to its prior odds.

Interpretation follows that $BF > 1$ means H_1 is more supported by the data than H_2.

Kass and Raftery [1995] pointed out that the BF is very general and does not require alternative models to be nested. Also from the definition, the BF embraces prior beliefs for evaluation so that it provides a way of incorporating external information about a hypothesis.

The calculation of the BF contains integrations, numerical methods are needed. Next chapter, we introduce one powerful method Markov chain Monte Carlo(MCMC) method.

Chapter 3

Markov chain Monte Carlo Method

Consider the problem of evaluating expectation like

$$E_{\pi}[T(X)] = \int T(x)\pi(x)dx.$$

In Bayesian inference, we are interested in posterior mean $E(\theta|y)$ or posterior variance $Var(\theta|y)$. Therefore, above problem is very important but it can be difficult to calculate.

One solution is to draw independent samples $(X^{(1)}, X^{(2)}, \cdots, X^{(N)})$ from $\pi(x)$, then we can approximate

$$E_{\pi}[T(X)] \approx \frac{1}{N}\sum_{t=1}^{N} T(X^{(t)})$$

According to the Law of large numbers, above approximation is adoptable. This method is Monte Carlo integration.

Furthermore, it is known that above approximation is still possible if we sample using a Markov chain. This is the main idea of MCMC method and there are two major approaches, Metropolis-Hasting algorithm and Gibbs sampler.

3.1 Metropolis-Hasting(MH) algorithm

In order to sample from the posterior distribution, we can do the following steps.

ALGORITHM

1. Choose transition(proposal) function $q(y|x)$
2. Initialize θ_0
3. For j from 1 to N

   3.1. Generate θ^* from $q(\theta|\theta_{j-1})$

   3.2. Calculate the importance ratio,

$$r = \frac{\pi(\theta^*)/q(\theta^*|\theta_{j-1})}{\pi(\theta_{j-1})/q(\theta_{j-1}|\theta^*)} = \frac{\pi(\theta^*)q(\theta_{j-1}|\theta^*)}{\pi(\theta_{j-1})q(\theta^*|\theta_{j-1})} \tag{3.1}$$

   3.3. Update

$$\theta_j = \begin{cases} \theta^* & \text{with probability } min(r, 1) \\ \theta_{j-1} & \text{otherwise} \end{cases} \tag{3.2}$$

Intuitively, it seems reasonable since 1) if the jump $\theta_{j-1} \rightarrow \theta^*$ increases the posterior density ($r > 1$) then $\theta_j = \theta^*$, or 2) θ_j remains with probability $(1 - r)$ to avoid being stuck in local modes.

Note that for $p(\theta|y)$, normalizing constant is not needed because it is canceled out when we calculate importance ratio. Therefore MH algorithm gives us a way to inference when posterior has no closed form.

Choice of Proposal Density

A common choice of proposal density is random walk proposal,

$$q(x|y) = f(|x - y|)$$

, then importance ratio r in (3.1) becomes

$$r = \frac{\pi(\theta^*)}{\pi(\theta_{j-1})}$$

since $q(x|y) = q(y|x)$. Possible choices of f include the multivariate normal density and the multivariate t density.

Another choice is independent proposal,

$$q(x|y) = f(x)$$

, f can be the multivariate normal or t density. The more similar f is to π, the better performance MH has.

Remark 3.1. *Gaussian random walk proposal can cause getting stuck in local modes, very slow convergence and also low acceptance rates. Lin et al. [1987] proposed employing a mixture of Gaussian proposal to overcome this problem. This approach makes the tails of proposal distribution thicker, enables good performance.*

$$\theta^*|\theta_{j-1}, k \quad \sim \quad N(\theta_{j-1}, k\Omega)$$

$$k = \begin{cases} 1 & w.p. \quad 0.85 \\ 9 & w.p. \quad 0.1 \\ 81 & w.p. \quad 0.05 \end{cases}$$

3.2 Gibbs sampler

For joint distribution $\pi(\theta, \phi)$, generating (θ, ϕ) jointly is difficult. In this situation, following sampling procedure can be applicable.

ALGORITHM

1. Initialize θ_0 and ϕ_0.
2. For j from 1 to N,

    2.1. Generate θ_j from $\pi(\theta|\phi_{j-1})$.
    2.2. Generate ϕ_j from $\pi(\phi|\theta_j)$.

MH within Gibbs sampler

We also can use both algorithms by applying MH algorithm inside the Gibbs sampler. For many parameters, Gibbs sampler gives a way to divide multidimensional problem into smaller dimensional problems, and MH gives a way to deal with normalizing constants.

For given $(\theta_1, \cdots, \theta_p)$, the strategy is to divide this vector into blocks. A general rule for blocking is to maximize within-block correlations and minimize the between-block correlations. For each block, we apply Gibbs sampler, and MH algorithm within blocks.

1. Initialize θ_0 and ϕ_0.
2. For j from 1 to N,

    2.1. Generate θ_j from $\pi(\theta|\phi_{j-1})$.

    a. Generate θ^* from proposal $q(\theta|\phi_{j-1}, \theta_{j-1})$
    b. Calculate importance rate r as in 3.1
    c. Set $\theta_j = \theta^*$ with probability $min(1, r)$

    2.2. Generate ϕ_j from $\pi(\phi|\theta_j)$.

    a. Generate ϕ^* from proposal $q(\phi|\phi_{j-1}, \theta_j)$
    b. Calculate importance rate r as in 3.1
    c. Set $\phi_j = \phi^*$ with probability $min(1, r)$

Remark 3.2. *The first m pre-chosen iterations of the MCMC sampling are discarded, and referred to as burn-in. This is used to avoid dependence of initial* *value and ensure that samples are indeed close enough to the samples from true distribution.*

Remark 3.3. *To check the convergence of MCMC iteration, mathematical approaches are difficult. Some plots are practically used such as autocorrelation function(ACF) plot, trace plot, and so on.*

- *Trace plot: The value of the drawn sample at each iteration versus the iteration number.*
- *ACF plot: Correlations between every drawn sample and it kth lag. Since our drawn samples form Markov chain, the ACF plot is expected decay exponentially as lag increased.*

Chapter 4

Application to Threshold Autoregressive Model with Stochastic Volatility

4.1 Introduction

In this chapter, we study examples of Bayesian financial time series analysis based on the former discussion. First, we solve an example problem of estimating threshold autoregressive model with stochastic volatility model, and investigate through a simulation. Then, we apply this model to a real data and carry out a hypothesis testing via using Bayes factor.

To begin with, we discuss the threshold autoregressive model and stochastic volatility model.

Threshold Autoregressive(TAR) model

Nonlinear models can explain various aspects of financial dynamics compared to linear models. In the class of these models, Threshold Autoregressive model(TAR) uses piecewise linear models to get a better approximation of the conditional mean structure, motivated by asymmetry in rising and decline pattern.

The results from Li and Lam [1995] also showed that the conditional mean structure could depend significantly on the rise and fall of the market in the previous day.

For time series y_t, it is said to follow TAR$(g; p_1, \cdots, p_g)$ with y_{t-d} as a threshold variable if

$$y_t = \phi_0^{(k)} + \sum_{i=1}^{p_k} \phi_i^{(k)} y_{t-i} + a_t^{(k)}, \quad r_{k-1} \leq y_{t-d} < r_k, \text{ for } k = 1, \cdots, g \tag{4.1}$$

where

g: number of regime,

$\{a_t^{(k)}\}$: innovation, i.i.d., $\sim N(0, \sigma_k^2)$

d: threshold lag, positive integer,

r_j: threshold variable, real, $-\infty = r_0 < r_1 < \cdots < r_g = \infty$

TAR model has not been widely used in practice because it is hard to estimate threshold values. Chen et al. [1995] proposed a procedure for estimating the threshold values and other parameters objectively via Bayesian inference with Gibbs sampler.

Stochastic Volatility(SV) model

Volatility is an important factor in financial or economic time series and has many applications, such as option trading, risk management, and so on.

One of approaches to model volatility is stochastic volatility model(SV) which introduce an innovation to the conditional variance equation of a_t.

For a_t, innovation or shock for time series y_t, it is said to follow SV model if

$$a_t = \sqrt{h_t}\epsilon_t, \quad logh_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \log h_{t-i} + \eta_t \tag{4.2}$$

where

ϵ_t: i.i.d. $\sim N(0,1)$

η_t : i.i.d. $\sim N(0,\sigma^2)$

$\{\epsilon_t\}$ and $\{\eta_t\}$ are independent.

Adding η_t, the innovation, considerably increase the flexibility of the model in describing the h_t compared to other volatility models. However, for each shock a_t the model uses two innovations, that makes it difficult to estimate SV model (Tsay [2010]). The MCMC method can be a solution for this.

4.2 Model

In this simulation we combine TAR (4.1) and SV (4.2). Consider following model.

$$\begin{aligned} y_t &= \begin{cases} \phi_0^{(1)} + \phi_1^{(1)} y_{t-1} + a_t &, \quad y_{t-d} \le r \\ \phi_0^{(2)} + \phi_1^{(2)} y_{t-1} + a_t &, \quad y_{t-d} > r \end{cases} \\ a_t &= \sqrt{h_t}\epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} N(0,1) \\ \log h_t &= \alpha_0 + \alpha_1 \log h_{t-1} + \eta_t, \quad \eta_t \overset{i.i.d.}{\sim} N(0,\sigma^2) \end{aligned} \tag{4.3}$$

where

> $\{\epsilon_t\}$ and $\{\eta_t\}$ are independent.
>
> d: delay lag
>
> r: threshold value

We are interested in estimating unknown parameter $\theta = (\boldsymbol{\phi}, \boldsymbol{\alpha}, \sigma^2, r, d)$ based on observation $\boldsymbol{y} = (y_1, \cdots, y_n)$. (where $\boldsymbol{\phi} = (\phi_0^{(1)}, \phi_1^{(1)}, \phi_0^{(2)}, \phi_1^{(2)})$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)$)

In this problem, maximum likelihood method is not applicable because of the existence of latent variables $\boldsymbol{h} = (h_1, \cdots, h_n)$. By using data augmentation(Tanner and Wong [1987]) in the Bayesian framework, we can overcome this difficulty.

4.3 Prior settings and sampling scheme

Applying data augmentation strategy, the parameter space is augmented to $(\theta, \boldsymbol{h})$. Conditioning on $\boldsymbol{h}$, likelihood $p(\boldsymbol{y}|\theta, \boldsymbol{h})$ has closed form.

Note that given $(\theta, \boldsymbol{h})$, the conditional likelihood is expressed as

$$p(\boldsymbol{y}|\theta, \boldsymbol{h}) = \prod_{t=s+1}^{T} \left[\sum_{j=1}^{2} \frac{1}{\sqrt{2\pi h_t}} \exp\left\{ -\frac{(y_t - \mu_t)^2}{h_t} \right\} I_{jt} \right] \tag{4.4}$$

where

$$\begin{aligned} \mu_t &= \phi_0^{(j)} + \phi_1^{(j)} y_{t-1} \\ I_{jt} &= I(r_{j-1} \leq y_{t-d} < r_j) \end{aligned}$$

If we assume independent priors, posterior density is generally given as multiplying (4.4) by prior density.

Specifically, we consider Gibbs sampling as following.

STEP1. Sample $\boldsymbol{h}$ from $f(\boldsymbol{h}|\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{\alpha}, \sigma^2, r, d)$

STEP2. Sample $\boldsymbol{\alpha}$ from $f(\boldsymbol{\alpha}|\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{h}, \sigma^2, r, d)$

STEP3. Sample σ^2 from $f(\sigma^2|\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{h}, \boldsymbol{\alpha}, r, d)$

STEP4. Sample $\boldsymbol{\phi}$ from $f(\boldsymbol{\phi}|\boldsymbol{y}, \boldsymbol{h}, \boldsymbol{\alpha}, \sigma^2, r, d)$

STEP5. Sample r from $f(r|\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{h}, \boldsymbol{\alpha}, \sigma^2, d)$

STEP6. Sample d from $f(d|\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{h}, \boldsymbol{\alpha}, \sigma^2, r)$

Inside each steps, we also use MH algorithm with mixed Gaussian proposal as in Remark 3.1.

STEP1. Volatility Vector $\boldsymbol{h}$

The volatility vector $\boldsymbol{h}$ is drawn element-wise. Jacquier et al. [1994] use following derivation of univariate conditional densities.

$$\begin{aligned}
& f(h_t|\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{h_{-t}}, \boldsymbol{\alpha}, \sigma^2, r, d) \\
& \quad \propto f(a_t|h_t, y_t, y_{t-1}, \phi) f(h_t|h_{t-1}, \alpha, \sigma^2) f(h_{t+1}|h_t, \alpha, \sigma^2) \\
& \quad \propto N\left(\frac{y_t - \mu_t}{\sqrt{h_t}}\right) N\left(\frac{\log h_t - \nu_t}{\sqrt{\sigma_h^2}}\right)
\end{aligned} \tag{4.5}$$

where $\boldsymbol{h_{-t}}$ is the vector of $\boldsymbol{h}$ excluded h_t, $N(\cdot)$ is the density of standard normal distribution, and

$$\begin{aligned}
\mu_t &= \begin{cases} \phi_0^{(1)} + \phi_1^{(1)} y_{t-1} \quad , \quad y_{t-d} \leq r \\ \phi_0^{(2)} + \phi_1^{(2)} y_{t-1} \quad , \quad y_{t-d} > r \end{cases} \\
\nu_t &= \alpha_0(1 - \alpha_1)/(1 + \alpha_1^2) + \alpha_1(\log h_{t-1} + \log h_{t+1})/(1 + \alpha_1^2) \\
\sigma_h^2 &= \sigma^2/(1 + \alpha_1^2)
\end{aligned}$$

,with our model assumption. Since normalizing constant is difficult to calculate, we use MH algorithm with (4.5).

Remark 4.1. *Another approach to sample volatility vector $\boldsymbol{h}$ is to use of the forward filtering and backward sampling within the Kalman filter framework. (Shephard [1994] and See Tsay [2010])*

STEP2. Volatility Coefficient $\boldsymbol{\alpha}$

Note that

$$f(\boldsymbol{\alpha}|\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{h}, \sigma^2, r, d) = f(\boldsymbol{\alpha}|\boldsymbol{h}, \sigma^2) \tag{4.6}$$

and the right-hand side of the above equation is the form of the AR(1) model since $\log h_t$ follows AR(1) given $\boldsymbol{h}$.

Therefore, we set the conjugate prior distribution of $\boldsymbol{\alpha}$ as multivariate normal

$$MVN(\boldsymbol{\alpha}_0, V_0)$$

, then the posterior distribution becomes

$$MVN(\boldsymbol{\alpha}_*, V_*)$$

where

$$\begin{aligned} \boldsymbol{\alpha}_* &= V_*(\sum_{t=2}^{n} z_t \log h_t/\sigma^2 + V_0^{-1}\boldsymbol{\alpha}_0), \\ V_*^{-1} &= \sum_{t=2}^{n} z_t z_t'/\sigma^2 + V_0^{-1} \\ z_t &= (1, \log h_{t-1})' \end{aligned}$$

STEP3. Volatility Innovation σ^2

Note that

$$f(\sigma^2|\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{h}, \boldsymbol{\alpha}, r, d) = f(\sigma^2|\boldsymbol{h}, \boldsymbol{\alpha}) \tag{4.7}$$

and the right-hand side of the above equation is the form of the AR(1) model.

We set the conjugate prior distribution as $(m\lambda)/\sigma^2 \sim \chi^2_m$, then

$$\frac{m\lambda + \sum_{t=2}^{n} \eta_t^2}{\sigma^2} \sim \chi^2_{m+n-1}$$

STEP4. TAR Coefficient $\boldsymbol{\phi}$

We set the prior distribution of $(\phi_0^{(k)}, \phi_1^{(k)})$ as

$$MVN(\boldsymbol{\phi}_{k0}, W_{k0})$$

and posterior distribution can be easily obtained similarly as **STEP2**.

STEP5. Threshold Variable r

We assume r follows uniform distribution on (l, u). l and u are suitably chosen as quantiles of the observation $\boldsymbol{y}$ to ensure sufficient sample size for valid inference (Chen et al. [1995]).

STEP6. Delay lag d

We assume d follows a discrete uniform distribution on $\{1, 2, \cdots, d_0\}$. Then posterior distribution is multinomial distribution with probability

$$p(d = i|\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{h}, \boldsymbol{\alpha}, \sigma^2, r) = \frac{f(\boldsymbol{\phi}, \boldsymbol{h}, \boldsymbol{\alpha}, \sigma^2, d = i, r|\boldsymbol{y})}{\sum_{j=1}^{d_0} f(\boldsymbol{\phi}, \boldsymbol{h}, \boldsymbol{\alpha}, \sigma^2, d = j, r|\boldsymbol{y})}$$

4.4 Simulation study

Until now we see TAR with SV model and Bayesian estimation methodologies. We now try simulation experiment through an example to investigate the result.

We consider the following model.

$$\begin{aligned}
y_t &= \begin{cases} 0.02 - 0.8y_{t-1} + a_t & , \quad y_{t-3} \le 0.5 \\ -3 + 0.5y_{t-1} + a_t & , \quad y_{t-3} > 0.5 \end{cases} \\
a_t &= \sqrt{h_t}\epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} N(0,1) \\
\log h_t &= -0.2 + 0.8\log h_{t-1} + \eta_t, \quad \eta_t \overset{i.i.d.}{\sim} N(0,1) \qquad (4.8)
\end{aligned}$$

The choice of autoregressive coefficients and stochastic volatility coefficients refer to the explanations of So et al. [2002] that the autoregressive coefficient is usually positive when $y_t \geq 0$ and the converse is true. Also we put the fact that high persistence in variance was discovered in most of the stochastic volatility literature.

T=500 samples are generated, 'R 3.0.2 for Windows' is used for simulation (some codes use Fortran for the speed issue) and Total 3000 iterations are conducted for sampling, and the first 1000 iterations are ignored as burn-in iterations.

Initial values are set to be $(0,0,0,0)$ for $\boldsymbol{\phi}$, $(0.1, 0.5)$ for $\boldsymbol{\alpha}$, 10 for σ^2, median of sample y_t for r, and 5 for d. Latent variable vector $\boldsymbol{h}$ are randomly chosen from $N(0,1)$.

Also hyperparameters -parameters in prior distribution- are set to be $\alpha_0 = (0,0)$, $V_0 = diag(10,10)$, $\phi_{k0} = (0,0)$, $W_{k0} = diag(10,10)$, l and u are 0.1, 0.9 quantiles of

y_t , $d_0 = 10$.

Fig.4.1 shows the generated sample y_t, h_t and 90% CI of sampled h_t.

Table 4.1 shows the result statistics of the estimated posterior distribution.

We can see that estimated values are similar to the true values and 95% confidence interval include the true values.

Table 4.1: Simulation results

Parameter	$\phi_0^{(1)}$	$\phi_1^{(1)}$	$\phi_0^{(2)}$	$\phi_1^{(2)}$	α_0	α_1	σ^2	r	d
True	0.02	-0.8	-3	0.5	-0.2	0.8	1	0.5	3
Mean	0.017	-0.775	-3.030	0.530	-0.172	0.812	1.021	0.456	3
Median	-0.022	-0.795	-3.016	0.519	-0.156	0.810	0.989	0.402	3
Std	0.047	0.029	0.051	0.038	0.068	0.038	0.137	0.113	0
95% CI low.	-0.040	-0.862	-3.069	0.440	-0.220	0.741	0.655	0.380	3
95% CI upp.	0.066	-0.746	-2.836	0.562	-0.078	0.893	1.261	0.614	3

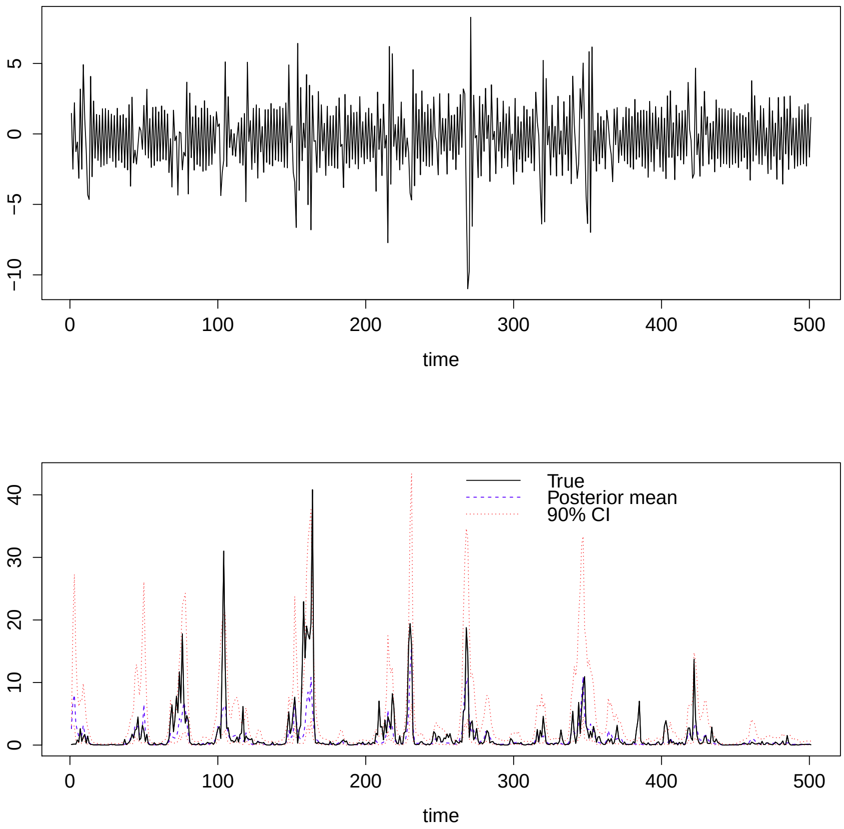

Figure 4.1: Generated time series y_t & h_t, and estimated volatility h_t

4.5 Empirical study

In this section, we see how to apply Bayesian inference to real data analysis. This section includes estimating parameters and hypothesis testing for model selection.

Chen et al. [1995] studied U.S. monthly civilian unemployment rates from the 2Q of 1948 to the 1Q of 1991 to estimate TAR(2;4,2) model. We use the same data but updated with 675 observations from 2Q of 1948 to the 1Q of 2004. This data is available in BAYSTAR package in R. Figure 4.2 shows the time plot of this data. Here we conduct analysis on this data set for the proposed model to compare result with only TAR model. We use the first differenced series $r_t = y_t - y_{t-1}$ since the sample autocorrelation function decays slowly.

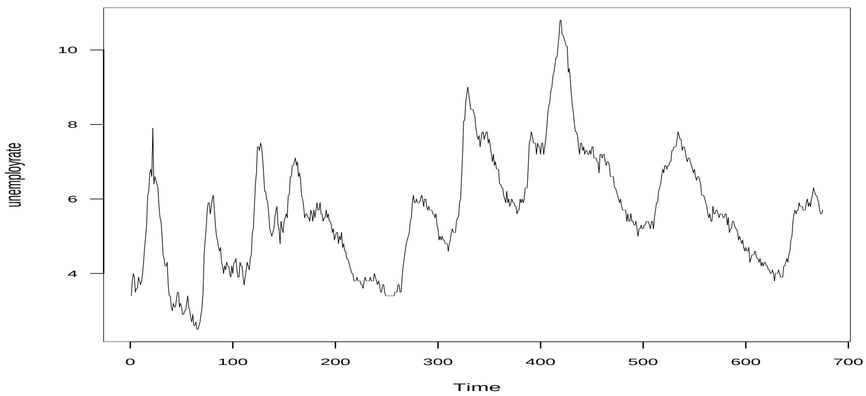

Figure 4.2: Time series plot of U.S. monthly unemployment rates

Estimation

Initial settings are similar to the former simulation study except that we use $AR(4)$ model for regime 1, and $AR(2)$ model for regime 2. The estimated parameters are in Table 4.2.

Table 4.2: Parameters estimated in real data analysis

Parameter	$\phi_0^{(1)}$	$\phi_1^{(1)}$	$\phi_2^{(1)}$	$\phi_3^{(1)}$	$\phi_4^{(1)}$	$\phi_0^{(2)}$	$\phi_1^{(2)}$	$\phi_2^{(2)}$	α_0	α_1	σ^2	r	d
Mean	0.003	-0.057	0.114	0.020	0.086	-0.032	0.227	0.255	-2.601	0.252	0.501	0.050	
Median	-0.003	-0.021	0.130	-0.008	0.134	-0.066	0.183	0.430	-2.541	0.243	0.502	0.050	2
Std	0.010	0.100	0.095	0.057	0.059	0.037	0.049	0.191	0.182	0.054	0.035	0.030	
95% CI low.	-0.003	-0.282	0.000	-0.008	0.000	-0.066	0.183	0.048	-3.117	0.119	0.446	0.003	2
95% CI upp.	0.025	0.000	0.285	0.147	0.134	0.009	0.280	0.430	-2.419	0.310	0.567	0.094	2

Compared to Chen et al. [1995], estimated coefficients in the mean structure, threshold variables are similar to them. With this model, we also can model the volatility of the series. To check the convergence, we practically use trace plot(Figure (4.3)) and autocorrelation plot(Figure (4.4)). From the figures, we can say that our MCMC samples are well converged to the samples from true posterior distribution.

Model Comparison

Next, we use Bayes factor to compare model. Assume that M_1 is our model, M_2 is the $TAR(2;4,2)$ model. If the calculated BF is greater than 1, we can conclude that our model is more supported by the data $\boldsymbol{y}$.

In the BF definition (2.1), it is hard to compute integrations. There are many numerical approximation methods for calculation of BF, here we use MCMC method with posterior density.

Note that in the definition of BF (Remark 4.2) ,

$$
\begin{aligned}
p(\boldsymbol{y}|M_j) &= \int p(\boldsymbol{y}|\Theta_j, M_j)\pi(\Theta_j|M_j)d\Theta_j \\
&\approx \left(\frac{1}{N-M} \sum_{i=1}^{N-M} p(\boldsymbol{y}|\Theta_j^{(i)}, M_j)^{-1} \right)^{-1}
\end{aligned}
\tag{4.9}
$$

where

$$
\begin{aligned}
\Theta_j &: \text{ the parameters of the Model } j \\
\pi(\Theta_j|M_j) &: \text{ the prior density under } M_j, \\
\Theta_j^{(i)} &: \text{ the sample drawn from } i\text{th MCMC iteration.}
\end{aligned}
$$

In this example, $\log BF = \log 1197.362 - \log 1000.282 = 0.180$ and $BF = 1.20$. Thus our model M_1 is more supported by the data.

Remark 4.2. *(4.9) is derived as below (see Kass and Raftery [1995])*

Dropping the notational dependence on M_j, then

$$p(y|M_j) \to p(y) = \int p(y|\theta)\pi(\theta)d\theta$$

The simplest Monte Carlo integration estimation is

$$\hat{p}(y|M_j) = \frac{1}{m}\sum_{i=1}^{m} p(y|\theta^{(i)})$$

where $\theta^{(i)}$ is the ith sample from prior distribution $\pi(\theta)$.

To improve estimation, above equation becomes,

$$\hat{p}(y|M_j) = \frac{\sum_{i=1}^{m} w_i p(y|\theta^{(i)})}{\sum_{i=1}^{m} w_i} \tag{4.10}$$

where

$w_i = \pi(\theta^{(i)})/p(\theta^{(i)}|y)$,

$\theta^{(i)}$ is the ith sample from posterior distribution $p(\theta|y)$.

Then use the following equation and substitute into the (4.10), these give the result of (4.9)

$$p(\theta|y) = \frac{p(y|\theta)\pi(\theta)}{p(y)}$$

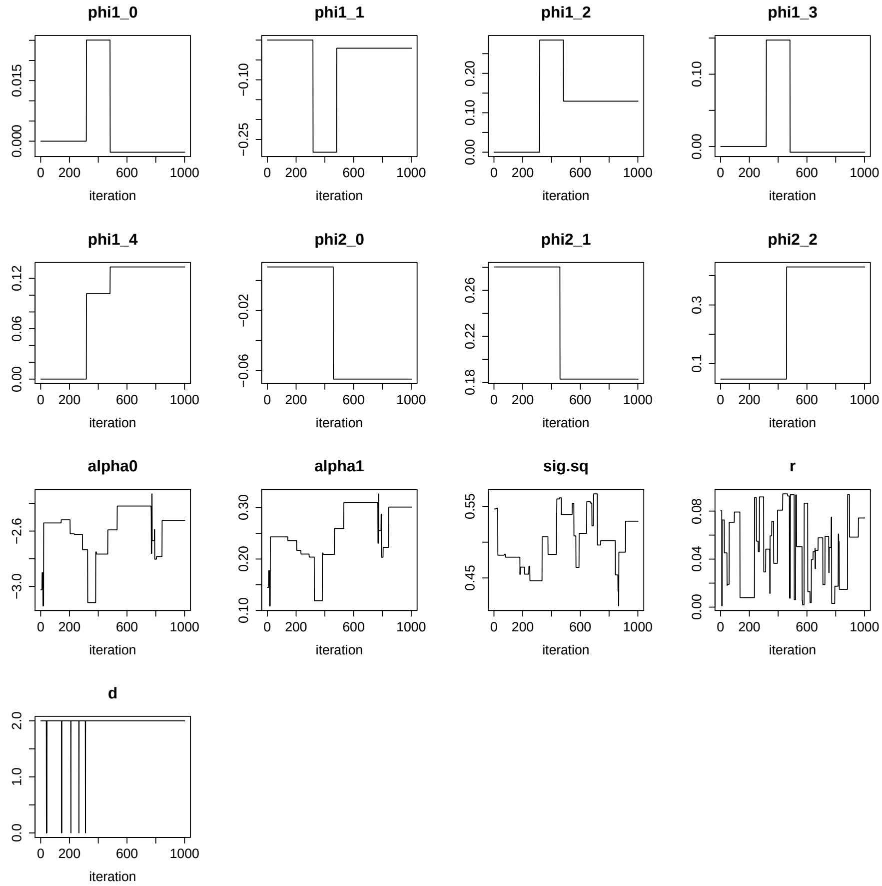

Figure 4.3: Trace plot of MCMC iteration in real data analysis

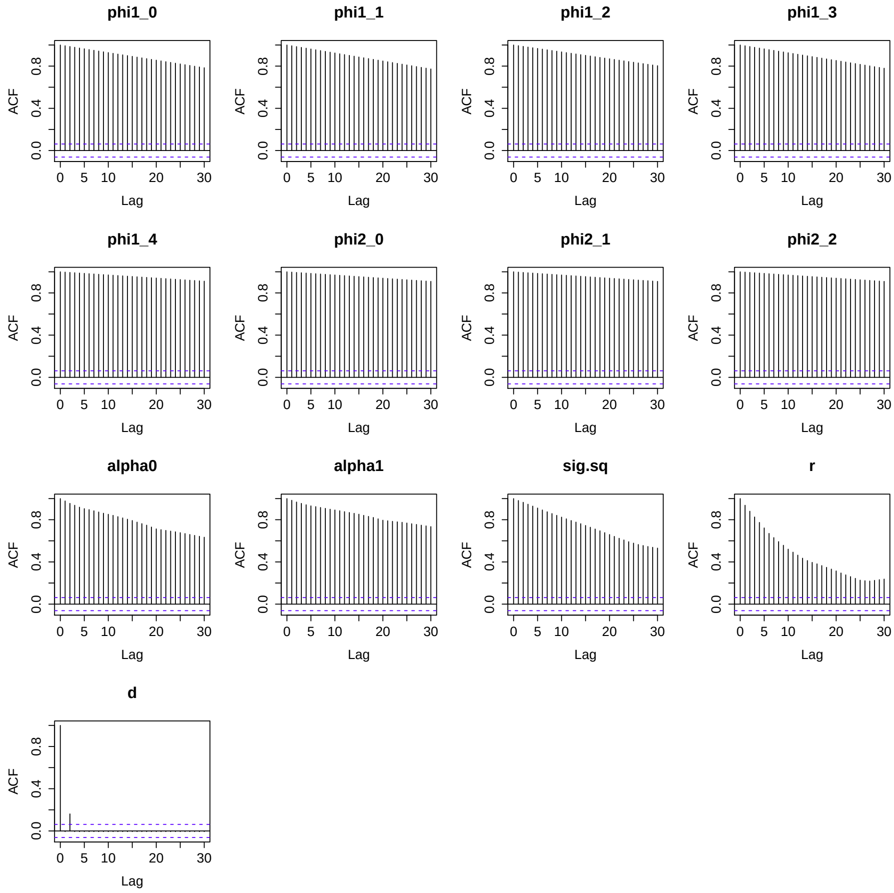

Figure 4.4: Autocorrelation function(ACF) plot of MCMC iteration in real data analysis

Chapter 5

Conclusion

We see the concepts and application of Bayesian financial time series via MCMC method. Unlike Frequentist approach, Bayesian inference is based on posterior distribution so that prior beliefs before observing data can be included in analysis. Also by Bayes factor, hypothesis testing can be flexible.

From an application example, we apply MCMC method to estimate threshold autoregressive model with stochastic volatility. Although TAR and SV model have very complex structure, by using MCMC we see that parameters can be nicely estimated. Also we studied real data to compare with other literature. After estimating parameters, we conduct hypothesis testing for model selection via Bayes factor. Here we see that by using BF, testing problem becomes relatively simple.

In addition to the problem of estimating parameters or model selection in this thesis, there are many other applications of Bayesian inference with MCMC. Applications include other model selection problem such as unit root test, also the important issue in financial time series. Therefore, the model we saw could be extended to unit root test, with an another approximate calculation of Bayes factor and mixed prior distributions proposed by Li and Yu [2010] and Chen et al. [2013].

Bibliography

Chen, C.W.S. and Lee, J.C. (1995) Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis* **16**, 483–492.

Chen, C.W.S., Chen, S.-Y. and Lee, S. (2013) Bayesian unit root test in double threshold heteroskedastic models. *Computational Economics* **42**, 471–490.

Jacquier, E., Polson, N. and Rossi, P.E. (1994) Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics* **12**, 371–417.

Kass, R. E., Raftery, A. E. (1995) Bayes factor. *Journal of the American Statistical Association* **90**, 773–795.

Li, W.K., Lam, K. (2010) Modelling asymmetry in stock returns by a threshold ARCH model. *The Statistician* **44**, 333–341.

Li, Y., Yu, J. (2010) A new bayesian unit root test in stochastic volatility models. *Research Collection School Of Economics.*

Lin, E.M.H., Chen, C.W.S., Gerlach, R. (2012) Forecasting volatility with asymmetric smooth transition dynamic range models. *International Journal of Forecasting* **28**, 384–399.

Prado, R., Wesk. M. (2010) *Time series: modeling, computation, and inference.* CRC Press, New York.

Shephard, N. (1994) Partial non-Gaussian state space. *Biometrika* **81**, 115–131.

So, M.K.P, Li, W.K. and Lam, K. (2002) On a threshold stochastic volatility model. *Journal of Forecasting* **21**, 473–500.

Tanner, M. A. and Wong, W.H. (1987) The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82**, 528–540.

Tsay, R. S. (2010) *Analysis of financial time Series, third Edition.* Wiley, New York.

Wasserman, L. (2004) *All of statistics.* Springer, New York.

국문 초록

본 논문에서는 금융 시계열 데이터의 베이지안 추론 방법에 대해서 소개하고 활용 예제를 제시한다. 베이지안 접근은 자료의 관측 전 모수에 대한 사전 확률분포를 이용하여 사후 분포를 계산하여 추론하는 방법으로서, 자료를 관측하기 전의 믿음을 분석에 활용할 수 있다는 장점이 있다. 또한 Markov chain Monte Carlo(MCMC) 등의 계산 방법을 통해 복잡한 모형을 비교적 쉽고 빠르게 분석할 수 있다.

따라서 본 논문에서는 먼저 전반적인 베이지안 추론 방법과 시계열 자료에 활용하는 방법에 대해서 살펴본 후, 계산을 위한 MCMC 방법에 대해서 소개한다. 또한 이러한 소개를 바탕으로 Stochastic Volatility를 갖는 Threshold Autoregressive 모형의 모수를 추정하는 문제를 시뮬레이션을 통해서 수행해보도록 한다. 이에 더해 위의 모형을 실제 시계열 자료에 적용해보도록 하고, 모형 선택 가설검정 예제를 Bayes factor를 이용한 방법을 통해 수행한다.

주요어 : 베이지안 시계열 분석, 혼합 사전 확률, 마르코프 연쇄 몬테칼로 방법, Threshold autoregressive, Stochastic volatility, Bayes factor.

학번 : 2012-20230

A Review on Bayesian Financial Time Series Analysis

지도교수 이 상 열

이 논문을 이학석사 학위논문으로 제출함

2014년 2월

서울대학교 대학원

통계학과

이 희 석

이희석의 이학석사 학위논문을 인준함

2014년 2월

위 원 장 ________________ (인)

부 위 원 장 ________________ (인)

위 원 ________________ (인)

이 학 석 사 학 위 논 문

A Review on Bayesian Finincial Time Series Analysis

금융 시계열 자료의 베이지안 분석에 대한 고찰

2014년 2월

서울대학교 대학원
통계학과
이 희 석

A Review on Bayesian Financial Time Series Analysis

by

Hee-seok Lee

A Thesis

submitted in fulfillment of the requirement

for the degree of

Master of Science

in Statistics

The Department of Statistics

College of Natural Sciences

Seoul National University

February, 2014

Abstract

This thesis introduces concepts and applications of Bayesian inference for financial time series. Bayesian inference is a method of statistical inference using Bayes' theorem to update prior beliefs as additional informations are observed. This allows us to use our prior beliefs of parameters and the Markov chain Monte Carlo method(MCMC) makes the analysis is relatively fast and simple.

In this thesis we introduce the time series Bayesian inference and the MCMC method, illustrate an example of estimating unknown parameters in threshold autoregressive(TAR) models with stochastic volatility(SV). Moreover, we apply TAR with SV model to a real data set and conduct a hypothesis test for model selection via using Bayes factor.

Key words : Bayesian time series, Mixed Prior, Markov chain Monte Carlo, Threshold autoregressive, Stochastic volatility, Bayes factor.

Student number : 2012-20230

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Advances of computational methodology have increased ability to solve complicated problem. They also extend the applicability of many existing statistical methods. Especially in the area of econometrics with many complex models, advantages of computational method have great importance.

Therefore, in this thesis we introduce one of the outstanding developments in computational methodology, Markov chain Monte Carlo(MCMC) method that are widely applicable in financial time series. First, in Chapter 2 we discuss Bayesian inference including general procedure and model comparison method via using Bayes factor. Bayesian inference provides us to insert prior beliefs of parameters before observing data. Second, in Chapter 3 we discuss algorithms of MCMC, which gives us the simple and fast way to calculate posterior density in Bayesian inference. Then, in Chapter 4 we focus on the financial time series problem. In particular, we carry out simulation and empirical studies based on former discussions.

For application, first we demonstrate simulation study of threshold autoregressive model(TAR) with stochastic volatility(SV). TAR model is one of tools to capture the nonlinearity of the financial time series. As well as mean structure, SV model is considered because modeling volatility has many applications in financial time series such as option trading or risk management. Also these two models are combined so this could be a good example of advantages of Bayesian inference via MCMC method in complicated financial time series problem. Second, we apply TAR with SV to a real data set and carry out a hypothesis testing via using Bayes factor for model selection problem.

Chapter 2

Bayesian Inference

Bayesian inference is a method of statistical inference by using Bayes' theorem to update prior beliefs when additional informations are observed. Bayesian inference is psychologically appealing because it allows us to insert our prior beliefs about parameters before data are observed.

Although it has a weakness of having subjective notion of probability, Bayesian approach has wide usage in many fields of statistics. In most cases solutions of two approaches are similar, even some cases Bayesian solutions might be advantageous.

2.1 General procedure

Bayesian inference is generally carried out in the following steps.

1. Choose the probability of parameter – prior distribution – before observing

data. Prior distribution reflects beliefs about parameter θ.

2. Choose model $f(x|\theta)$ reflects beliefs about x given θ.

3. After seeing data $X_1, \cdots, X_n$, use Bayes' theorem to calculate the posterior distribution $f(\theta|X_1, \cdots, X_n)$.

For the time series analysis, steps are similar.

Suppose we observe time series data $\boldsymbol{y} = (y_1, \cdots, y_n)$ from $\{y_t; t \geq 0\}$, collection of random variables over time. If we believe that y_t has a some density function $p(\cdot|\theta)$, our observation can be written as $p(\boldsymbol{y}|\theta)$. When we see this as the function of θ, we call it the likelihood function. Unlike Frequentist approaches that are mostly based on this likelihood function, Bayesian introduce pre-assumed beliefs called 'prior', $\pi(\theta)$

By Bayes' theorem,

$$\begin{aligned} p(\theta|y) &= \frac{p(y|\theta)\pi(\theta)}{\int_\Theta p(y|\theta)\pi(\theta)d\theta} \\ p(\theta|y) &\propto p(y|\theta)\pi(\theta) \end{aligned}$$

$p(\cdot|\theta)$ and $\pi(\cdot)$ denote likelihood and prior density function, respectively. Also we call $p(\theta|y)$ as posterior density. Bayesian inferences are based on this posterior distribution.

From the choice of prior distribution, it is classified into conjugate and nonconjugate analysis.

Conjugate Bayesian analysis

For convenience, the prior distribution often assumed to be conjugate, which is from same distribution family with corresponding posterior distribution. Advantages are its reasonable features and the simple calculation that may result in closed analytical form.

Nonconjugate Bayesian analysis

However, in many situations, there is no closed analytical form of posterior distribution against our desire.

It is hard to calculate integration in the denominator of posterior density function. It may need numerical approximation, or other methods.

Since the Markov chain Monte Carlo(MCMC) method introduced, it become possible to inference when we have no closed form of posterior distribution by obtaining sample draws from it. Next chapter, we see the algorithms of the MCMC method.

2.2 Bayes factor

For the model comparison or hypothesis testing, Bayesian approach uses Bayes factor(BF), the Bayesian version of the classical likelihood ratio test(LRT).

Consider two models or hypotheses H_1 and H_2 for given data y, Bayes factor is defined as,

$$BF = \frac{p(y|H_1)}{p(y|H_2)} = \frac{\int p(\theta_1|H_1)p(y|\theta_1, H_1)d\theta_1}{\int p(\theta_2|H_2)p(y|\theta_2, H_2)d\theta_2} \tag{2.1}$$

where θ_i s stand for parameters in H_i s.

Note Bayes' theorem says $p(H_j|y) \propto p(y|H_j)p(H_j)$. Therefore we can see that *posterior odds ratio* $\propto$ *BF* $\times$ *prior odds ratio.* The BF can be translated as ratio of the posterior odds to its prior odds.

Interpretation follows that $BF > 1$ means H_1 is more supported by the data than H_2.

Kass and Raftery [1995] pointed out that the BF is very general and does not require alternative models to be nested. Also from the definition, the BF embraces prior beliefs for evaluation so that it provides a way of incorporating external information about a hypothesis.

The calculation of the BF contains integrations, numerical methods are needed. Next chapter, we introduce one powerful method Markov chain Monte Carlo(MCMC) method.

Chapter 3

Markov chain Monte Carlo Method

Consider the problem of evaluating expectation like

$$E_\pi[T(X)] = \int T(x)\pi(x)dx.$$

In Bayesian inference, we are interested in posterior mean $E(\theta|y)$ or posterior variance $Var(\theta|y)$. Therefore, above problem is very important but it can be difficult to calculate.

One solution is to draw independent samples $(X^{(1)}, X^{(2)}, \cdots, X^{(N)})$ from $\pi(x)$, then we can approximate

$$E_\pi[T(X)] \approx \frac{1}{N}\sum_{t=1}^{N} T(X^{(t)})$$

According to the Law of large numbers, above approximation is adoptable. This method is Monte Carlo integration.

Furthermore, it is known that above approximation is still possible if we sample using a Markov chain. This is the main idea of MCMC method and there are two major approaches, Metropolis-Hasting algorithm and Gibbs sampler.

3.1 Metropolis-Hasting(MH) algorithm

In order to sample from the posterior distribution, we can do the following steps.

ALGORITHM

1. Choose transition(proposal) function $q(y|x)$
2. Initialize θ_0
3. For j from 1 to N

   3.1. Generate θ^* from $q(\theta|\theta_{j-1})$

   3.2. Calculate the importance ratio,

$$r = \frac{\pi(\theta^*)/q(\theta^*|\theta_{j-1})}{\pi(\theta_{j-1})/q(\theta_{j-1}|\theta^*)} = \frac{\pi(\theta^*)q(\theta_{j-1}|\theta^*)}{\pi(\theta_{j-1})q(\theta^*|\theta_{j-1})} \tag{3.1}$$

   3.3. Update

$$\theta_j = \begin{cases} \theta^* & \text{with probability } min(r,1) \\ \theta_{j-1} & \text{otherwise} \end{cases} \tag{3.2}$$

Intuitively, it seems reasonable since 1) if the jump $\theta_{j-1} \to \theta^*$ increases the posterior density ($r > 1$) then $\theta_j = \theta^*$, or 2) θ_j remains with probability $(1 - r)$ to avoid being stuck in local modes.

Note that for $p(\theta|y)$, normalizing constant is not needed because it is canceled out when we calculate importance ratio. Therefore MH algorithm gives us a way to inference when posterior has no closed form.

Choice of Proposal Density

A common choice of proposal density is random walk proposal,

$$q(x|y) = f(|x - y|)$$

, then importance ratio r in (3.1) becomes

$$r = \frac{\pi(\theta^*)}{\pi(\theta_{j-1})}$$

since $q(x|y) = q(y|x)$. Possible choices of f include the multivariate normal density and the multivariate t density.

Another choice is independent proposal,

$$q(x|y) = f(x)$$

, f can be the multivariate normal or t density. The more similar f is to π, the better performance MH has.

Remark 3.1. *Gaussian random walk proposal can cause getting stuck in local modes, very slow convergence and also low acceptance rates. Lin et al. [1987] proposed employing a mixture of Gaussian proposal to overcome this problem. This approach makes the tails of proposal distribution thicker, enables good performance.*

$$\theta^*|\theta_{j-1}, k \sim N(\theta_{j-1}, k\Omega)$$

$$k = \begin{cases} 1 & w.p. \quad 0.85 \\ 9 & w.p. \quad 0.1 \\ 81 & w.p. \quad 0.05 \end{cases}$$

3.2 Gibbs sampler

For joint distribution $\pi(\theta, \phi)$, generating (θ, ϕ) jointly is difficult. In this situation, following sampling procedure can be applicable.

ALGORITHM

1. Initialize θ_0 and ϕ_0.
2. For j from 1 to N,

    2.1. Generate θ_j from $\pi(\theta|\phi_{j-1})$.
    2.2. Generate ϕ_j from $\pi(\phi|\theta_j)$.

MH within Gibbs sampler

We also can use both algorithms by applying MH algorithm inside the Gibbs sampler. For many parameters, Gibbs sampler gives a way to divide multidimensional problem into smaller dimensional problems, and MH gives a way to deal with normalizing constants.

For given $(\theta_1, \cdots, \theta_p)$, the strategy is to divide this vector into blocks. A general rule for blocking is to maximize within-block correlations and minimize the between-block correlations. For each block, we apply Gibbs sampler, and MH algorithm within blocks.

1. Initialize θ_0 and ϕ_0.
2. For j from 1 to N,

   2.1. Generate θ_j from $\pi(\theta|\phi_{j-1})$.

      a. Generate θ^* from proposal $q(\theta|\phi_{j-1}, \theta_{j-1})$
      b. Calculate importance rate r as in 3.1
      c. Set $\theta_j = \theta^*$ with probability $min(1, r)$

   2.2. Generate ϕ_j from $\pi(\phi|\theta_j)$.

      a. Generate ϕ^* from proposal $q(\phi|\phi_{j-1}, \theta_j)$
      b. Calculate importance rate r as in 3.1
      c. Set $\phi_j = \phi^*$ with probability $min(1, r)$

Remark 3.2. *The first m pre-chosen iterations of the MCMC sampling are discarded, and referred to as burn-in. This is used to avoid dependence of initial* *value and ensure that samples are indeed close enough to the samples from true distribution.*

Remark 3.3. *To check the convergence of MCMC iteration, mathematical approaches are difficult. Some plots are practically used such as autocorrelation function(ACF) plot, trace plot, and so on.*

- *Trace plot: The value of the drawn sample at each iteration versus the iteration number.*

- *ACF plot: Correlations between every drawn sample and it kth lag. Since our drawn samples form Markov chain, the ACF plot is expected decay exponentially as lag increased.*

Chapter 4

Application to Threshold Autoregressive Model with Stochastic Volatility

4.1 Introduction

In this chapter, we study examples of Bayesian financial time series analysis based on the former discussion. First, we solve an example problem of estimating threshold autoregressive model with stochastic volatility model, and investigate through a simulation. Then, we apply this model to a real data and carry out a hypothesis testing via using Bayes factor.

To begin with, we discuss the threshold autoregressive model and stochastic volatility model.

Threshold Autoregressive(TAR) model

Nonlinear models can explain various aspects of financial dynamics compared to linear models. In the class of these models, Threshold Autoregressive model(TAR) uses piecewise linear models to get a better approximation of the conditional mean structure, motivated by asymmetry in rising and decline pattern.

The results from Li and Lam [1995] also showed that the conditional mean structure could depend significantly on the rise and fall of the market in the previous day.

For time series y_t, it is said to follow TAR$(g; p_1, \cdots, p_g)$ with y_{t-d} as a threshold variable if

$$y_t = \phi_0^{(k)} + \sum_{i=1}^{p_k} \phi_i^{(k)} y_{t-i} + a_t^{(k)}, \quad r_{k-1} \le y_{t-d} < r_k, \text{ for } k = 1, \cdots, g \tag{4.1}$$

where

g: number of regime,

$\{a_t^{(k)}\}$: innovation, i.i.d., $\sim N(0, \sigma_k^2)$

d: threshold lag, positive integer,

r_j: threshold variable, real, $-\infty = r_0 < r_1 < \cdots < r_g = \infty$

TAR model has not been widely used in practice because it is hard to estimate threshold values. Chen et al. [1995] proposed a procedure for estimating the threshold values and other parameters objectively via Bayesian inference with Gibbs sampler.

Stochastic Volatility(SV) model

Volatility is an important factor in financial or economic time series and has many applications, such as option trading, risk management, and so on.

One of approaches to model volatility is stochastic volatility model(SV) which introduce an innovation to the conditional variance equation of a_t.

For a_t, innovation or shock for time series y_t, it is said to follow SV model if

$$a_t = \sqrt{h_t}\epsilon_t, \quad logh_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \log h_{t-i} + \eta_t \tag{4.2}$$

where

ϵ_t: i.i.d. $\sim N(0,1)$

η_t : i.i.d. $\sim N(0,\sigma^2)$

$\{\epsilon_t\}$ and $\{\eta_t\}$ are independent.

Adding η_t, the innovation, considerably increase the flexibility of the model in describing the h_t compared to other volatility models. However, for each shock a_t the model uses two innovations, that makes it difficult to estimate SV model (Tsay [2010]). The MCMC method can be a solution for this.

4.2 Model

In this simulation we combine TAR (4.1) and SV (4.2). Consider following model.

$$\begin{aligned}
y_t &= \begin{cases} \phi_0^{(1)} + \phi_1^{(1)} y_{t-1} + a_t & , \quad y_{t-d} \le r \\ \phi_0^{(2)} + \phi_1^{(2)} y_{t-1} + a_t & , \quad y_{t-d} > r \end{cases} \\
a_t &= \sqrt{h_t}\epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} N(0,1) \\
\log h_t &= \alpha_0 + \alpha_1 \log h_{t-1} + \eta_t, \quad \eta_t \overset{i.i.d.}{\sim} N(0,\sigma^2) \qquad (4.3)
\end{aligned}$$

where

$\{\epsilon_t\}$ and $\{\eta_t\}$ are independent.

d: delay lag

r: threshold value

We are interested in estimating unknown parameter $\theta = (\boldsymbol{\phi}, \boldsymbol{\alpha}, \sigma^2, r, d)$ based on observation $\boldsymbol{y} = (y_1, \cdots, y_n)$. (where $\boldsymbol{\phi} = (\phi_0^{(1)}, \phi_1^{(1)}, \phi_0^{(2)}, \phi_1^{(2)})$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)$)

In this problem, maximum likelihood method is not applicable because of the existence of latent variables $\boldsymbol{h} = (h_1, \cdots, h_n)$. By using data augmentation(Tanner and Wong [1987]) in the Bayesian framework, we can overcome this difficulty.

4.3 Prior settings and sampling scheme

Applying data augmentation strategy, the parameter space is augmented to $(\theta, \boldsymbol{h})$. Conditioning on $\boldsymbol{h}$, likelihood $p(\boldsymbol{y}|\theta, \boldsymbol{h})$ has closed form.

Note that given $(\theta, \boldsymbol{h})$, the conditional likelihood is expressed as

$$p(\boldsymbol{y}|\theta, \boldsymbol{h}) = \prod_{t=s+1}^{T} \left[\sum_{j=1}^{2} \frac{1}{\sqrt{2\pi h_t}} \exp\left\{ -\frac{(y_t - \mu_t)^2}{h_t} \right\} I_{jt} \right] \tag{4.4}$$

where

$$\begin{aligned} \mu_t &= \phi_0^{(j)} + \phi_1^{(j)} y_{t-1} \\ I_{jt} &= I(r_{j-1} \leq y_{t-d} < r_j) \end{aligned}$$

If we assume independent priors, posterior density is generally given as multiplying (4.4) by prior density.

Specifically, we consider Gibbs sampling as following.

STEP1. Sample $\boldsymbol{h}$ from $f(\boldsymbol{h}|\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{\alpha}, \sigma^2, r, d)$

STEP2. Sample $\boldsymbol{\alpha}$ from $f(\boldsymbol{\alpha}|\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{h}, \sigma^2, r, d)$

STEP3. Sample σ^2 from $f(\sigma^2|\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{h}, \boldsymbol{\alpha}, r, d)$

STEP4. Sample $\boldsymbol{\phi}$ from $f(\boldsymbol{\phi}|\boldsymbol{y}, \boldsymbol{h}, \boldsymbol{\alpha}, \sigma^2, r, d)$

STEP5. Sample r from $f(r|\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{h}, \boldsymbol{\alpha}, \sigma^2, d)$

STEP6. Sample d from $f(d|\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{h}, \boldsymbol{\alpha}, \sigma^2, r)$

Inside each steps, we also use MH algorithm with mixed Gaussian proposal as in Remark 3.1.

STEP1. Volatility Vector $\boldsymbol{h}$

The volatility vector $\boldsymbol{h}$ is drawn element-wise. Jacquier et al. [1994] use following derivation of univariate conditional densities.

$$\begin{aligned} & f(h_t|\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{h_{-t}}, \boldsymbol{\alpha}, \sigma^2, r, d) \\ & \quad \propto f(a_t|h_t, y_t, y_{t-1}, \phi) f(h_t|h_{t-1}, \alpha, \sigma^2) f(h_{t+1}|h_t, \alpha, \sigma^2) \\ & \quad \propto N\left(\frac{y_t - \mu_t}{\sqrt{h_t}}\right) N\left(\frac{\log h_t - \nu_t}{\sqrt{\sigma_h^2}}\right) \end{aligned} \tag{4.5}$$

where $\boldsymbol{h_{-t}}$ is the vector of $\boldsymbol{h}$ excluded h_t, $N(\cdot)$ is the density of standard normal distribution, and

$$\begin{aligned} \mu_t &= \begin{cases} \phi_0^{(1)} + \phi_1^{(1)} y_{t-1} & , \quad y_{t-d} \le r \\ \phi_0^{(2)} + \phi_1^{(2)} y_{t-1} & , \quad y_{t-d} > r \end{cases} \\ \nu_t &= \alpha_0(1-\alpha_1)/(1+\alpha_1^2) + \alpha_1(\log h_{t-1} + \log h_{t+1})/(1+\alpha_1^2) \\ \sigma_h^2 &= \sigma^2/(1+\alpha_1^2) \end{aligned}$$

,with our model assumption. Since normalizing constant is difficult to calculate, we use MH algorithm with (4.5).

Remark 4.1. *Another approach to sample volatility vector $\boldsymbol{h}$ is to use of the forward filtering and backward sampling within the Kalman filter framework. (Shephard [1994] and See Tsay [2010])*

STEP2. Volatility Coefficient $\boldsymbol{\alpha}$

Note that

$$f(\boldsymbol{\alpha}|\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{h}, \sigma^2, r, d) = f(\boldsymbol{\alpha}|\boldsymbol{h}, \sigma^2) \tag{4.6}$$

and the right-hand side of the above equation is the form of the AR(1) model since $\log h_t$ follows AR(1) given $\boldsymbol{h}$.

Therefore, we set the conjugate prior distribution of $\boldsymbol{\alpha}$ as multivariate normal

$$MVN(\boldsymbol{\alpha}_0, V_0)$$

, then the posterior distribution becomes

$$MVN(\boldsymbol{\alpha}_*, V_*)$$

where

$$\begin{aligned} \boldsymbol{\alpha}_* &= V_*(\sum_{t=2}^{n} z_t \log h_t/\sigma^2 + V_0^{-1}\boldsymbol{\alpha}_0), \\ V_*^{-1} &= \sum_{t=2}^{n} z_t z_t'/\sigma^2 + V_0^{-1} \\ z_t &= (1, \log h_{t-1})' \end{aligned}$$

STEP3. Volatility Innovation σ^2

Note that

$$f(\sigma^2|\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{h}, \boldsymbol{\alpha}, r, d) = f(\sigma^2|\boldsymbol{h}, \boldsymbol{\alpha}) \tag{4.7}$$

and the right-hand side of the above equation is the form of the AR(1) model.

We set the conjugate prior distribution as $(m\lambda)/\sigma^2 \sim \chi^2_m$, then

$$\frac{m\lambda + \sum_{t=2}^{n} \eta_t^2}{\sigma^2} \sim \chi^2_{m+n-1}$$

STEP4. TAR Coefficient $\boldsymbol{\phi}$

We set the prior distribution of $(\phi_0^{(k)}, \phi_1^{(k)})$ as

$$MVN(\boldsymbol{\phi}_{k0}, W_{k0})$$

and posterior distribution can be easily obtained similarly as **STEP2**.

STEP5. Threshold Variable r

We assume r follows uniform distribution on (l, u). l and u are suitably chosen as quantiles of the observation $\boldsymbol{y}$ to ensure sufficient sample size for valid inference (Chen et al. [1995]).

STEP6. Delay lag d

We assume d follows a discrete uniform distribution on $\{1, 2, \cdots, d_0\}$. Then posterior distribution is multinomial distribution with probability

$$p(d = i|\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{h}, \boldsymbol{\alpha}, \sigma^2, r) = \frac{f(\boldsymbol{\phi}, \boldsymbol{h}, \boldsymbol{\alpha}, \sigma^2, d = i, r|\boldsymbol{y})}{\sum_{j=1}^{d_0} f(\boldsymbol{\phi}, \boldsymbol{h}, \boldsymbol{\alpha}, \sigma^2, d = j, r|\boldsymbol{y})}$$

4.4 Simulation study

Until now we see TAR with SV model and Bayesian estimation methodologies. We now try simulation experiment through an example to investigate the result.

We consider the following model.

$$
\begin{aligned}
y_t &= \begin{cases} 0.02 - 0.8y_{t-1} + a_t & , \quad y_{t-3} \le 0.5 \\ -3 + 0.5y_{t-1} + a_t & , \quad y_{t-3} > 0.5 \end{cases} \\
a_t &= \sqrt{h_t}\epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} N(0,1) \\
\log h_t &= -0.2 + 0.8\log h_{t-1} + \eta_t, \quad \eta_t \overset{i.i.d.}{\sim} N(0,1)
\end{aligned} \tag{4.8}
$$

The choice of autoregressive coefficients and stochastic volatility coefficients refer to the explanations of So et al. [2002] that the autoregressive coefficient is usually positive when $y_t \ge 0$ and the converse is true. Also we put the fact that high persistence in variance was discovered in most of the stochastic volatility literature.

T=500 samples are generated, 'R 3.0.2 for Windows' is used for simulation (some codes use Fortran for the speed issue) and Total 3000 iterations are conducted for sampling, and the first 1000 iterations are ignored as burn-in iterations.

Initial values are set to be $(0,0,0,0)$ for $\boldsymbol{\phi}$, $(0.1, 0.5)$ for $\boldsymbol{\alpha}$, 10 for σ^2, median of sample y_t for r, and 5 for d. Latent variable vector $\boldsymbol{h}$ are randomly chosen from $N(0,1)$.

Also hyperparameters -parameters in prior distribution- are set to be $\alpha_0 = (0,0)$, $V_0 = diag(10,10)$, $\phi_{k0} = (0,0)$, $W_{k0} = diag(10,10)$, l and u are 0.1, 0.9 quantiles of

y_t , $d_0 = 10$.

Fig.4.1 shows the generated sample y_t, h_t and 90% CI of sampled h_t.

Table 4.1 shows the result statistics of the estimated posterior distribution.

We can see that estimated values are similar to the true values and 95% confidence interval include the true values.

Table 4.1: Simulation results

Parameter	$\phi_0^{(1)}$	$\phi_1^{(1)}$	$\phi_0^{(2)}$	$\phi_1^{(2)}$	α_0	α_1	σ^2	r	d
True	0.02	-0.8	-3	0.5	-0.2	0.8	1	0.5	3
Mean	0.017	-0.775	-3.030	0.530	-0.172	0.812	1.021	0.456	3
Median	-0.022	-0.795	-3.016	0.519	-0.156	0.810	0.989	0.402	3
Std	0.047	0.029	0.051	0.038	0.068	0.038	0.137	0.113	0
95% CI low.	-0.040	-0.862	-3.069	0.440	-0.220	0.741	0.655	0.380	3
95% CI upp.	0.066	-0.746	-2.836	0.562	-0.078	0.893	1.261	0.614	3

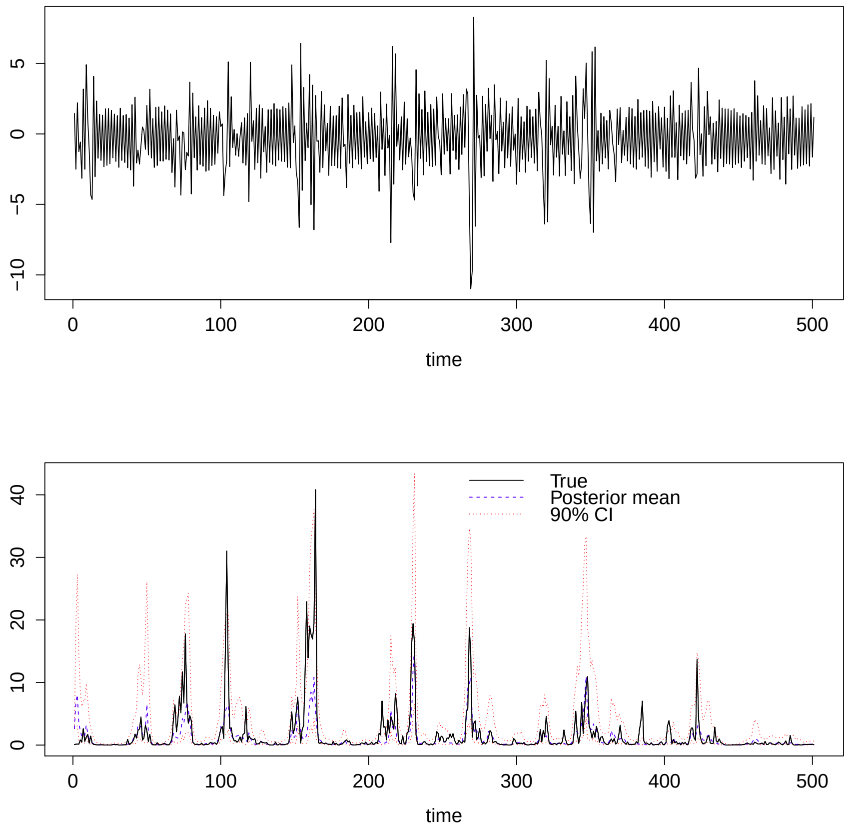

Figure 4.1: Generated time series y_t & h_t, and estimated volatility h_t

4.5 Empirical study

In this section, we see how to apply Bayesian inference to real data analysis. This section includes estimating parameters and hypothesis testing for model selection.

Chen et al. [1995] studied U.S. monthly civilian unemployment rates from the 2Q of 1948 to the 1Q of 1991 to estimate TAR(2;4,2) model. We use the same data but updated with 675 observations from 2Q of 1948 to the 1Q of 2004. This data is available in BAYSTAR package in R. Figure 4.2 shows the time plot of this data. Here we conduct analysis on this data set for the proposed model to compare result with only TAR model. We use the first differenced series $r_t = y_t - y_{t-1}$ since the sample autocorrelation function decays slowly.

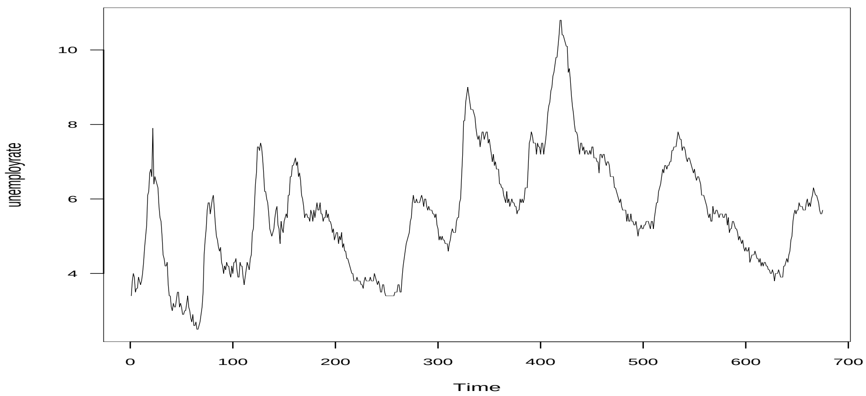

Figure 4.2: Time series plot of U.S. monthly unemployment rates

Estimation

Initial settings are similar to the former simulation study except that we use $AR(4)$ model for regime 1, and $AR(2)$ model for regime 2. The estimated parameters are in Table 4.2.

Table 4.2: Parameters estimated in real data analysis

Parameter	$\phi_0^{(1)}$	$\phi_1^{(1)}$	$\phi_2^{(1)}$	$\phi_3^{(1)}$	$\phi_4^{(1)}$	$\phi_0^{(2)}$	$\phi_1^{(2)}$	$\phi_2^{(2)}$	α_0	α_1	σ^2	r	d
Mean	0.003	-0.057	0.114	0.020	0.086	-0.032	0.227	0.255	-2.601	0.252	0.501	0.050	
Median	-0.003	-0.021	0.130	-0.008	0.134	-0.066	0.183	0.430	-2.541	0.243	0.502	0.050	2
Std	0.010	0.100	0.095	0.057	0.059	0.037	0.049	0.191	0.182	0.054	0.035	0.030	
95% CI low.	-0.003	-0.282	0.000	-0.008	0.000	-0.066	0.183	0.048	-3.117	0.119	0.446	0.003	2
95% CI upp.	0.025	0.000	0.285	0.147	0.134	0.009	0.280	0.430	-2.419	0.310	0.567	0.094	2

Compared to Chen et al. [1995], estimated coefficients in the mean structure, threshold variables are similar to them. With this model, we also can model the volatility of the series. To check the convergence, we practically use trace plot(Figure (4.3)) and autocorrelation plot(Figure (4.4)). From the figures, we can say that our MCMC samples are well converged to the samples from true posterior distribution.

Model Comparison

Next, we use Bayes factor to compare model. Assume that M_1 is our model, M_2 is the $TAR(2;4,2)$ model. If the calculated BF is greater than 1, we can conclude that our model is more supported by the data $\boldsymbol{y}$.

In the BF definition (2.1), it is hard to compute integrations. There are many numerical approximation methods for calculation of BF, here we use MCMC method with posterior density.

Note that in the definition of BF (Remark 4.2) ,

$$\begin{aligned} p(\boldsymbol{y}|M_j) &= \int p(\boldsymbol{y}|\Theta_j, M_j)\pi(\Theta_j|M_j)d\Theta_j \\ &\approx \left(\frac{1}{N-M} \sum_{i=1}^{N-M} p(\boldsymbol{y}|\Theta_j^{(i)}, M_j)^{-1} \right)^{-1} \end{aligned} \tag{4.9}$$

where

$$\begin{aligned} \Theta_j &: \text{ the parameters of the Model } j \\ \pi(\Theta_j|M_j) &: \text{ the prior density under } M_j, \\ \Theta_j^{(i)} &: \text{ the sample drawn from } i\text{th MCMC iteration.} \end{aligned}$$

In this example, $\log BF = \log 1197.362 - \log 1000.282 = 0.180$ and $BF = 1.20$. Thus our model M_1 is more supported by the data.

Remark 4.2. *(4.9) is derived as below (see Kass and Raftery [1995])*

Dropping the notational dependence on M_j, then

$$p(y|M_j) \to p(y) = \int p(y|\theta)\pi(\theta)d\theta$$

The simplest Monte Carlo integration estimation is

$$\hat{p}(y|M_j) = \frac{1}{m}\sum_{i=1}^{m} p(y|\theta^{(i)})$$

where $\theta^{(i)}$ is the ith sample from prior distribution $\pi(\theta)$.

To improve estimation, above equation becomes,

$$\hat{p}(y|M_j) = \frac{\sum_{i=1}^{m} w_i p(y|\theta^{(i)})}{\sum_{i=1}^{m} w_i} \tag{4.10}$$

where

$w_i = \pi(\theta^{(i)})/p(\theta^{(i)}|y)$,

$\theta^{(i)}$ is the ith sample from posterior distribution $p(\theta|y)$.

Then use the following equation and substitute into the (4.10), these give the result of (4.9)

$$p(\theta|y) = \frac{p(y|\theta)\pi(\theta)}{p(y)}$$

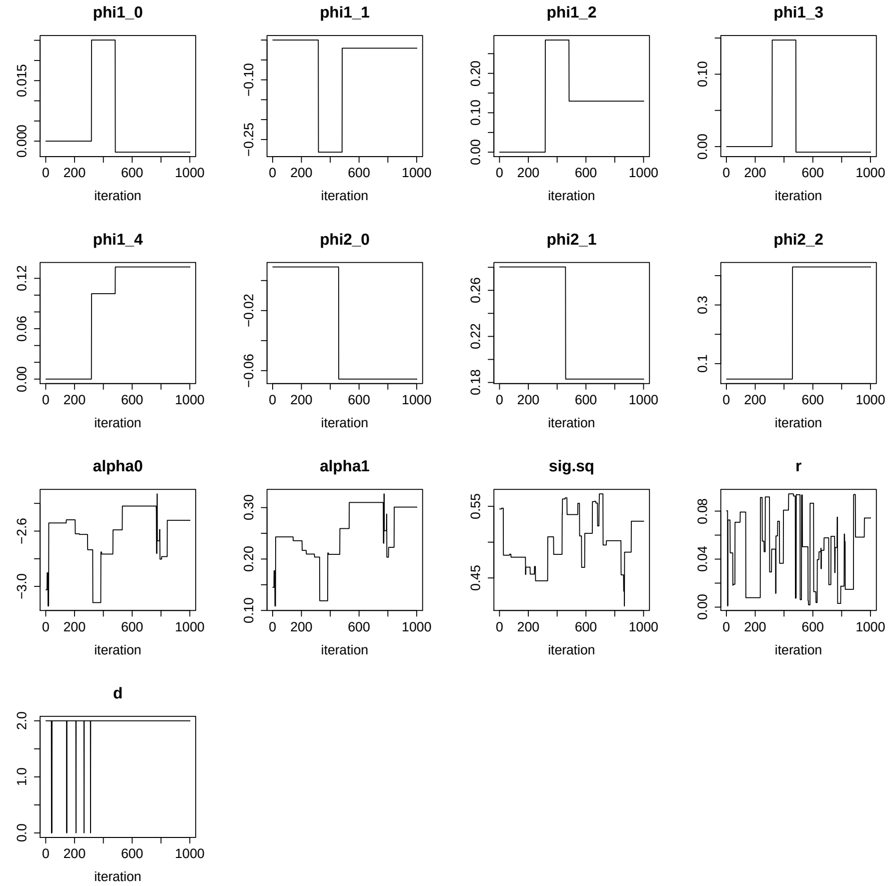

Figure 4.3: Trace plot of MCMC iteration in real data analysis

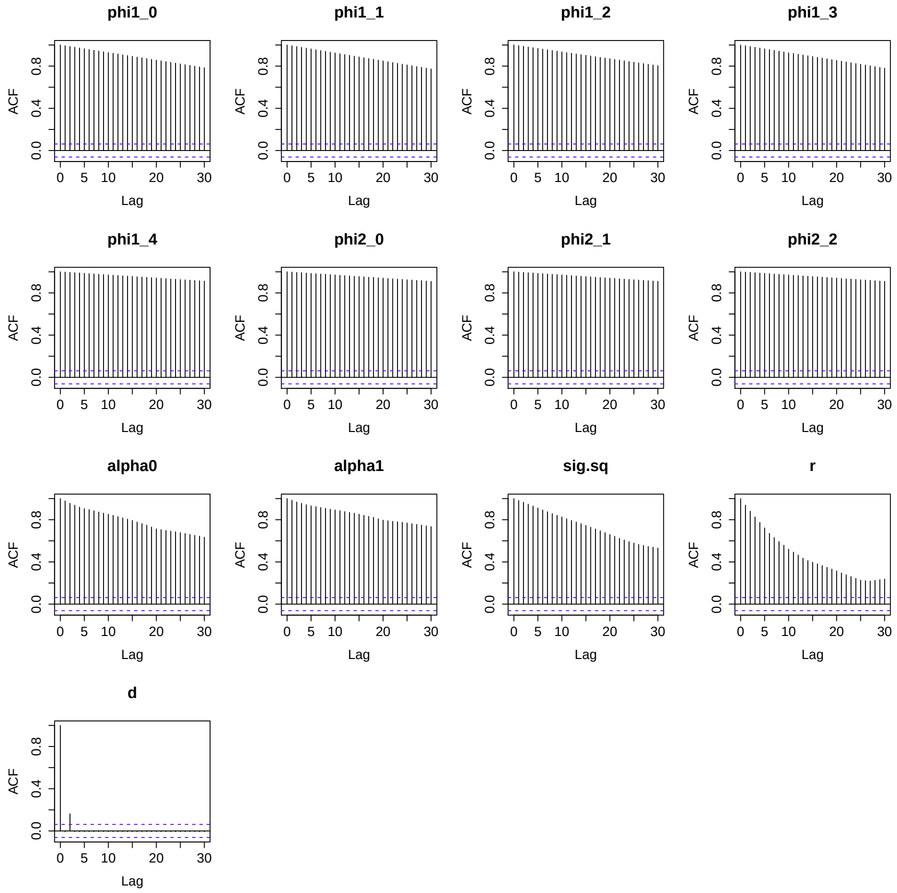

Figure 4.4: Autocorrelation function(ACF) plot of MCMC iteration in real data analysis

Chapter 5

Conclusion

We see the concepts and application of Bayesian financial time series via MCMC method. Unlike Frequentist approach, Bayesian inference is based on posterior distribution so that prior beliefs before observing data can be included in analysis. Also by Bayes factor, hypothesis testing can be flexible.

From an application example, we apply MCMC method to estimate threshold autoregressive model with stochastic volatility. Although TAR and SV model have very complex structure, by using MCMC we see that parameters can be nicely estimated. Also we studied real data to compare with other literature. After estimating parameters, we conduct hypothesis testing for model selection via Bayes factor. Here we see that by using BF, testing problem becomes relatively simple.

In addition to the problem of estimating parameters or model selection in this thesis, there are many other applications of Bayesian inference with MCMC. Applications include other model selection problem such as unit root test, also the important issue in financial time series. Therefore, the model we saw could be extended to unit root test, with an another approximate calculation of Bayes factor and mixed prior distributions proposed by Li and Yu [2010] and Chen et al. [2013].

Bibliography

Chen, C.W.S. and Lee, J.C. (1995) Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis* **16**, 483–492.

Chen, C.W.S., Chen, S.-Y. and Lee, S. (2013) Bayesian unit root test in double threshold heteroskedastic models. *Computational Economics* **42**, 471–490.

Jacquier, E., Polson, N. and Rossi, P.E. (1994) Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics* **12**, 371–417.

Kass, R. E., Raftery, A. E. (1995) Bayes factor. *Journal of the American Statistical Association* **90**, 773–795.

Li, W.K., Lam, K. (2010) Modelling asymmetry in stock returns by a threshold ARCH model. *The Statistician* **44**, 333–341.

Li, Y., Yu, J. (2010) A new bayesian unit root test in stochastic volatility models. *Research Collection School Of Economics.*

Lin, E.M.H., Chen, C.W.S., Gerlach, R. (2012) Forecasting volatility with asymmetric smooth transition dynamic range models. *International Journal of Forecasting* **28**, 384–399.

Prado, R., Wesk. M. (2010) *Time series: modeling, computation, and inference.* CRC Press, New York.

Shephard, N. (1994) Partial non-Gaussian state space. *Biometrika* **81**, 115–131.

So, M.K.P, Li, W.K. and Lam, K. (2002) On a threshold stochastic volatility model. *Journal of Forecasting* **21**, 473–500.

Tanner, M. A. and Wong, W.H. (1987) The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82**, 528–540.

Tsay, R. S. (2010) *Analysis of financial time Series, third Edition.* Wiley, New York.

Wasserman, L. (2004) *All of statistics.* Springer, New York.

국문 초록

본 논문에서는 금융 시계열 데이터의 베이지안 추론 방법에 대해서 소개하고 활용 예제를 제시한다. 베이지안 접근은 자료의 관측 전 모수에 대한 사전 확률분포를 이용하여 사후 분포를 계산하여 추론하는 방법으로서, 자료를 관측하기 전의 믿음을 분석에 활용할 수 있다는 장점이 있다. 또한 Markov chain Monte Carlo(MCMC) 등의 계산 방법을 통해 복잡한 모형을 비교적 쉽고 빠르게 분석할 수 있다.

따라서 본 논문에서는 먼저 전반적인 베이지안 추론 방법과 시계열 자료에 활용하는 방법에 대해서 살펴본 후, 계산을 위한 MCMC 방법에 대해서 소개한다. 또한 이러한 소개를 바탕으로 Stochastic Volatility를 갖는 Threshold Autoregressive 모형의 모수를 추정하는 문제를 시뮬레이션을 통해서 수행해보도록 한다. 이에 더해 위의 모형을 실제 시계열 자료에 적용해보도록 하고, 모형 선택 가설검정 예제를 Bayes factor를 이용한 방법을 통해 수행한다.

주요어 : 베이지안 시계열 분석, 혼합 사전 확률, 마르코프 연쇄 몬테칼로 방법, Threshold autoregressive, Stochastic volatility, Bayes factor.

학번 : 2012-20230