BAYESIAN TIME SERIES: FINANCIAL MODELS AND SPECTRAL ANALYSIS

by

Yang Chen

Department of Statistics and Decision Sciences
Duke University

Date: _____

Approved:

_____
Dr. Mike West, Supervisor

_____
Dr. Peter Müller

_____
Dr. Brani Vidakovic

_____
Dr. S. Viswanathan

_____

Dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Department of Statistics and Decision Sciences
in the Graduate School of
Duke University

1997

Statistics

ABSTRACT

()

BAYESIAN TIME SERIES: FINANCIAL MODELS AND SPECTRAL ANALYSIS

by

Yang Chen

Department of Statistics and Decision Sciences
Duke University

Date: _____

Approved:

_____
Dr. Mike West, Supervisor

_____
Dr. Peter Müller

_____
Dr. Brani Vidakovic

_____
Dr. S. Viswanathan

_____

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Statistics and Decision Sciences in the Graduate School of
Duke University

1997

Abstract

This dissertation studies two models in Bayesian time series analysis: the Stochastic Volatility Model in the time domain and the Harmonic Model in the frequency domain of time series.

Volatility plays a central role in modern finance especially in the pricing of derivative securities. Research on changing volatility can be categorized into two groups: the time-varying volatility models represented by ARCH type models and the Stochastic Volatility Models. Research on ARCH type models offers straightforward implementation and has been empirically successful but they generally lack economic intuition. Stochastic Volatility models are statistically elegant and have strong connection to continuous-time finance models. Yet estimation of Stochastic Volatility models has been very difficult which makes further development of the model and comparison of empirical results with ARCH/GARCH type models difficult.

We propose an efficient Bayesian Markov Chain Monte Carlo estimation procedure for a Log-AR(1) Stochastic Volatility Model. We develop new simulation-based model diagnostics methods for in-sample model adequacy check and compare with the popular EGARCH model. Our in-sample diagnostic check shows better kurtosis properties and different Smile effect generated by the Stochastic Volatility Model than the EGARCH model. We also discuss issues of the comparison of historical volatility and implied volatility and propose a new model which combines the historical volatility and implied volatility under one model framework. This new model can be used for both forecasting and testing of the hypothesis of the existence of stochastic volatility.

Two common methods exist for frequency estimation in cyclical time series: prob-

ability theory and Fourier transform. Recent work of Jaynes and Bretthorst has shown the connection of the two methods and the theoretical advantage of the probability method. We develop a unified approach for accurate frequency estimation under the Bayesian MCMC framework for the single-frequency and multi-frequency harmonic model which can be generalized to more complex models for the frequency. We apply the method to real cyclical data. Motivated by the study of Oxygen isotope data in geology study, we discuss timing issues in harmonic analysis, particularly the impact of uncertain timing to the estimation of frequencies. We develop a harmonic model with uncertain timing to investigate the impact of uncertain timing in frequency estimation and to illustrate the use of Bayesian MCMC simulation methods as a general method for complex models in Bayesian spectral analysis. We illustrate our idea using real Oxygen data in geology study and provide evidence of the impact of uncertain timing to frequency estimation.

Acknowledgements

I would like to express a deep feeling of gratitude and appreciation to my advisor, Professor Mike West. His insights into various aspects of my thesis and patient mentoring have made writing this dissertation one of the most valuable experiences in my life. I am very grateful for having this opportunity and privilege to work with him. His constant challenge and encouragement and his attitudes towards research and life has greatly influenced and enriched my life.

I would like to thank Professor Peter Müller and Professor Brani Vidakovic for taking an interest in my work and offering valuable suggestions for improvement. I would like to specially thank Professor S. Viswanathan from the Fuqua School of Business at Duke for introducing me to the field of financial statistics. His patient guidance and the many hours of discussions he spent with me has made my study and research in this area a very enjoyable process.

I would like to thank Mr. Mike Henderson, Dr. Cedomir Crnkovic and Dr. Jordan Drachman at the proprietary trading group of J.P. Morgan for providing the data used in this dissertation and for countless hours of discussions on the first part of this dissertation.

I thank the faculty and my fellow graduate students at ISDS for providing such a stimulating and cooperative environment of research and study. I thank my friends from various departments at Duke for many stimulating talks which truly open my eyes to many other scientific fields and for sharing the feelings of life as a Ph.D. student.

Last but not least, I thank my family for their constant support and understanding of me, without which life would have been much harder.

Contents

Abstract **iv**

Acknowledgements **vi**

List of Figures **x**

List of Tables **xiv**

1 Introduction **1**

2 Log-AR(1) Stochastic Volatility Model **5**

2.1 The Model 5

2.2 Parameter Estimation 7

2.2.1 Classical Estimation Methods 7

2.2.2 A Single-move Gibbs Sampler 10

2.2.3 A Multi-move Gibbs Sampler 14

2.3 Application to Exchange Rate Data 17

2.4 Model Diagnostics 20

2.5 Conclusion 25

3 Model Extension **27**

3.1 A Log-AR(p) Stochastic Volatility Model 27

3.2 Correlated Log-AR(1) Stochastic Volatility Model 33

3.2.1 A Correlated Stochastic Volatility Model 33

3.2.2 A Threshold Stochastic Volatility Model 35

3.3 Combining Historical and Implied Volatility in one Model 37

3.3.1 The Model 39

3.3.2 MCMC for the Model 42
3.3.3 Model Testing 46
3.3.4 Extension to the Log-AR(1) Model 47

4 Model Comparison **50**

4.1 The Smile Effect 51
4.2 Comparing Financial Forecasting Models 53
4.3 Bayesian Option Pricing with Stochastic Volatility 58

5 Conclusions and Future directions **61**

6 Introduction **66**

7 A Single-frequency Harmonic Model **69**

7.1 The Model 70
7.2 Posterior Distribution for the Frequency 71
7.3 Metropolis Sampling of the Posterior Distribution 73
7.3.1 Mode-Based Proposal Distribution 75
7.3.2 Kernel Density Proposal Distribution 76
7.4 Estimation of Other Model Parameters 78
7.5 Application 79

8 A Multi-frequency Harmonic Model **89**

8.1 The Model 89
8.2 Parameter Estimation 90
8.2.1 Sampling of the Joint Posterior 92
8.2.2 Sample $\{\omega_1, \omega_2\}$ iteratively 95
8.2.3 Sample $\{\omega_1, \omega_2\}$ conditioning on other parameters 96

8.3 Application 97
8.3.1 EEG Data 97
8.3.2 Oxygen Data 99

9 Timing Uncertainty **106**

9.1 Timing Issues in Time Series Studies 107
9.2 Timing Uncertainty in Harmonic Models 110
9.3 An Harmonic Model with Uncertain Time 116
9.4 Application 121
9.5 Conclusion 128

10 Conclusions and Future directions **133**

Bibliography **136**

Biography **144**

List of Figures

2.1 *The log$\chi^2(1)$ distribution and its normal approximation* 9

2.2 *Left: Trajectory plots, acf and histogram of b for a Single-move Gibbs Sampler with $b = 0.95$, $\sigma = 0.1$. Right: comparison with $b = 0.5$, $\sigma = 0.3$.* 13

2.3 *Trajectory plot, acf and histogram of b for a Multi-move Gibbs Sampler with $b = 0.95$, $\sigma = 0.1$* 18

2.4 *Comparison of trajectory plot, acf and histogram of b for the DM/Dollar data using both the Single-move and Multi-move Gibbs Sampler* 20

2.5 *Histogram of the skewness and kurtosis statistics for the DEM/USD Data* 22

2.6 *Histogram of acf of the residuals of lag 1 to 9 for the DM/Dollar Data* 23

2.7 *QQ-plot of the sequence of normal variables from the predictive residual diagnostic of the DM/Dollar Data* 25

3.1 *Simulation trajactory, histogram and acf of b_1 and b_2 for the Log-AR(2) model with $b_1 = 0.65$, $b_2 = 0.2$* 31

3.2 *Simulation trajectory, histogram and acf of b_1 and b_2 for DM/Dollar data with the log-AR(2) model* 32

4.1 *Histogram of the posterior distribution of the last day's volatility* . . . 51

4.2 *Payoff of a straddle* 55

4.3 *Calculation of delta* 57

7.1 *Time series plot of the EEG data. The acf, periodogram , Schuster's periodogram, posterior distribution and log-posterior distribution of the frequency.* 80

7.2 *Histogram of parameter estimates for a,b,v and λ. With $p(\lambda|Y)$ on top of the histogram.* 82

7.3 *Top: EEG series and fitted values. Bottom: residuals* 83

7.4 *$p(\omega|Y)$ from different segment of the data indicates time-varying frequencies* 84

7.5 *Top: Residual plot of the EEG data after taking out 8 frequencies. Bottom: acf plot of the residuals* 85

7.6 *Graphical summary of the output produced by CODA plots* 86

7.7 *Geweke Convergence Diagnostic* 87

8.1 *Histogram of model parameters for two-frequency model on EEG data* 99

8.2 *Histogram of the four wavelengths of the EEG data* 100

8.3 *Histogram of the amplitudes for the four-frequency model* 101

8.4 *Time series plot of the Oxygen data* 102

8.5 *Posterior $p(\lambda|Y)$ for all 866 Oxygen data and for the first and second half of the Oxygen data* 103

8.6 *Histogram of model parameters for the two-frequency model of the first half of the Oxygen data* 104

8.7 *Histogram of the four wavelengths for the four-frequency model of the first half of the Oxygen data* 105

9.1 *Original depth of the core measurement versus $\delta^{18}O$ measurement series* 109

9.2 *The first 400 oxygen data under unequally spaced time 3y scale vs. under equally spaced time scale* 110

9.3 *Log-posterior for the first 400 Oxygen data under equally spaced time scale compared with that under unequally spaced time scale* 111

9.4 *Log-posterior under equally spaced time scale compared with that from a uncertain time scale with error from a Normal distribution with a s.d.=0.017* 112

9.5 *Log-posterior under original equally spaced time scale compared with that from simulated time from* $N(0, 0.33)$ 113

9.6 *Log-posterior under original time scale and simulated time from* $U(t_i - 0.3, t_i + 0.3)$ 114

9.7 *Log-posterior under original time scale and simulated time from* $U(s_{i-1}, s_{i+1})$ 116

9.8 *An example of true time vs. equally spaced time* 119

9.9 *Histograms of the opsteriors for wavelengths using a four-frequency model without timing error* 122

9.10 *Trajectory plots of simulations of the wavelengths under independent timing error from* $U(s_i - 0.3, s_i + 0.3)$ 123

9.11 *Posterior distributions of the wavelengths using a four-frequency model with independent timing error* 124

9.12 *Top: Time series plot of true time minus equally spaced time. Bottom: Time series plot of the difference of adjacent true time.* 125

9.13 *Trajectory plots of simulations of the wavelengths under dependent timing error* 126

9.14 *ACF of the samples from the oxygen data with dependent prior* 127

9.15 *Histogram of the wavelengths for the Oxygen data under dependent timing error* 128

9.16 *Top: Time series plot of true time minus equally spaced time. Bottom: Time series plot of the difference of adjacent true time.* 129

9.17 *Posterior distribution for the wavelength for all 300, the first 150 and the last 150 EEG data* 130

9.18 *Histogram of the model parameters for the EEG data under equally spaced time* 130

9.19 *Simulation trajectory plots for* λ, a, b, v 131

9.20 *Simulation trajectory plots for time* $t = 50$,$t = 100$, $t = 150$, $t = 200$, $t = 250$ *and* $t = 300$ 131

9.21 *Histogram of wavelength and other model parameters* 132

9.22 *Top: Time series plot of true time minus equally spaced time. Bottom: Time series plot of the difference of adjacent true time.* 132

List of Tables

2.1 Mean and Variance of the Mixture of Normals and the mixing probability 15

2.2 Data: Parameter Estimates for Single-move and Multi-move Gibbs Sampler 19

2.3 Statistics from the predictive residual diagnostic 25

3.1 Examples of Continuous-Time Stochastic Volatility Processes 48

4.1 Payoff of a straddle 56

7.1 Statistics from the predictive residual diagnostic 86

9.1 Means and variances of the four wavelengths under no timing error, independent timing error and dependent timing error 125

Chapter 1

Introduction

Volatility plays a central role in modern finance especially in the pricing of derivative securities. The Black-Scholes model for the pricing of an European option is by far the most widely used option pricing formula. Yet it is well known that the assumption of constant volatility of the Black-Scholes model is violated in the market. Research on changing volatility using time series models has been active in the last twenty years. There are two major types of volatility models: the determinstic time-varying volatility models represented by ARCH type models and the stochastic time-varying volatility models represented by Stochastic Volatility models.

ARCH/GARCH type models were pioneered by Engle (1982) and grew rapidly into a rich family of empirical models for volatility forecasting during the 1980's. They model the volatility as a linear function of the square of past observations and are therefore also called observation-driven models (Shephard, N. 1996). Such representation provides explicit one-step-ahead forecasting distribution for the volatility and offers straightforward implementation and interpretation. Empirical results indicated ARCH/GARCH type models have been very successful in forecasting volatilities and are popular among practitioners. They also parallel directly with the successful autoregressive and moving average models. The major criticism on ARCH/GARCH type models were their lack of economic intuition. Surveys of work on ARCH/GARCH type models are given in the papers by Bollerslev, T. Chou, R. and Kroner, K. (1992), Bollerslev, Engle and Nelson (1994) and Engle (1995).

Stochastic Volatility Models have long attracted researchers in finance and econometricians since the 1970's for various theorectical reasons. One use of the Stochastic

Volatility Models is to explain the random behavior of financial markets and theories on option pricing. Early work on using Stochastic Volatility Models to explain the random behavior of the market includes Clark (1973) and Tauchen and Pitts (1983) and Gallant,Hsieh and Tauchen (1991). Clark used stochastic volatility to represent the random and uneven flow of new information to the financial market. Tauchen and Pitts (1983) and Gallant, Hsieh and Tauchen(1991) refined this work and proposed a mixture of distributions of asset returns with temporal dependence in information arrivals. Another popular area of research on Stochastic Volatility Models is option pricing theory using Stochastic Volatility models. These include Hull and White (1987) who suggested a diffusion process for asset prices with volatility following a positive diffusion process, Melino and Turnbull (1991) who used a log-autoregressive stochastic process for the volatility, and Johnson and Shanno (1987), Scott (1987), Stein and Stein (1991), among others.

There are two reasons for the concentration of theoretical research on Stochastic Volatility models: modern derivative pricing theory is based on continuous-time stochastic processes with which Stochastic Volatility Models has a close resemblance, and estimation of stochastic volatility models has been a very difficult task. Unlike ARCH type models which model volatility as a function of past volatility, they introduce a seperate stochastic process for the volatility and are therefore also called parameter-driven models. Their statistical properties are easier to understand and generalize. Unfortunately, estimating the model is very difficult because of the non-linear structure of the model. The lack of efficient estimation procedures and therefore empirical results seriously hampered theoretical research on Stochastic Volatility Models.

Recent developments in estimating non-linear latent variable models has made efficient estimation of Stochastic Volatility Models a reality. This makes extensive

research on studying the empirical aspect of Stochastic Volatility Models feasible. In this part of the dissertation, we modify the pioneer work of Jacquier, Polson and Rossi (1994) and develop a more efficient Bayesian Markov Chain Monte Carlo estimation procedure for a Log-AR(1) Stochastic Volatility Model. We also develop new simulation based model diagnostics methods. We then compare some aspects of the Stochastic Volatility Model with the popular EGARCH model in the ARCH/GARCH model family using real exchange rate data. Finally, we discuss issues of the comparison of historical volatility and implied volatility and propose a new model which combines both the historical volatility and implied volatility under one model framework. The techniques developed on parameter estimation, residual diagnostic and model comparison can be generalized to other non-linear time series models.

This part of the dissertation is organized as follows: Chapter 2 develops a complete Bayesian analysis on the Log-AR(1) Stochastic Volatility Model. This includes an efficient MCMC estimation procedure whose performance is compared with the MCMC estimation procedure of Jacquier, Polson and Rossi (1994), new model dignostic procedures for the Log-AR(1) Stochastic Volatility Model, and illustrations of the methods on real exchange rate data. Chapter 3 develops three extensions to the basic Log-AR(1) Stochastic Volatility Model developed in Chapter 2. These include the Log-AR(p) Stochastic Volatility Model, the correlated Log-AR(1) Stochastic Volatility Model and the model which combines both historical volatility and implied volatility. Chapter 4 discusses the issue of model comparison under the context of financial forecasting models. In particular, an option trading strategy was developed for the comparison of the Log-AR(1) Stochastic Volatility Model and the EGARCH model. Chapter 5 discusses future directions on the study of Stochastic Volatility Models.

Chapter 2

Log-AR(1) Stochastic Volatility Model

Motivated by the work of Jacquier, Polson and Rossi (1994), we propose an efficient Bayesian Markov Chain Monte Carlo (MCMC) simulation method in this chapter to estimate a Log-AR(1) Stochastic Volatility Model. We first discuss the model and traditional estimation techniques for the model. We then propose two different MCMC techniques (a Single-Move and a Multi-Move Gibbs Sampler) to estimate the model and compare their performance. We apply the model to real exchange rate data and study the convergence behavior of both samplers. We then develop model diagnostic tools to check model adequacy on real exchange rate data.

2.1 The Model

A good forecasting model should capture all the forecastable patterns in the data and generate similar random behavior as observed in the market. To compare two forecasting models statistically, we need to compare both their in-sample performance and out-of-sample performance. Models with both good in-sample and out-of-sample performance will engender confidence in application to real market.

Before we proceed, we list some intuitive reasons for the use of Stochastic Volatility Model to forecast volatility. Among other empirical findings, the following stylized facts about volatilities are commonly observed in financial data: volatility clustering; fat tails distribution of asset returns; and the leverage effect (asymmetrical effects of price change on volatility). A good volatility forecasting model should capture all

these patterns.

There are many versions of Stochastic Volatility Models in the literature, varying through the choices of stochastic processes used to characterize the change of volatility. In virtually all current published continuous-time models, the stochastic process representing the underlying volatility is represented by a Wiener process. We adopt a discretized Log-AR(1) Stochastic Volatility Model proposed by Melino-Turnbull (1987) to model foreign exchange rates.

$$y_t = e^{h_t/2}\epsilon_t, \tag{2.1}$$

$$h_{t+1} = \alpha + \beta h_t + \sigma_h \mu_t. \tag{2.2}$$

Here y_t is the return of the asset and h_t is the log-volatility at time t. The log-volatility process respects the non-negativity of volatility and follows an AR(1) process. Both ϵ_t and μ_t are independent standard normal errors.

This model explains the stylized facts of volatility in a very elegant way. The Log-AR(1) process for the volatility captures volatility clustering, high values of b (close to 1) indicates high degree of volatility clustering; the additional random variance in the volatility process generates excess kurtosis in the marginal distribution of returns which leads to the fat tail distribution of the data; if we allow the two random noise ϵ_t and μ_t be correlated, this model neatly generates the asymmetrical effects of price change to volatility. In this chapter, however, we will assume ϵ_t and μ_t be independent.

2.2 Parameter Estimation

Although intuitively sound and statistically elegant, practical uses of Stochastic Volatility Models have been limited because of difficulties in model fitting. The

derivation of the likelihood function of model parameters $\omega = \{a, b, \sigma_h\}$ involves an N dimensional integration problem where the latent volatility process is being integrated out (N as the number of observations). Such computation becomes prohibitive with large N. This is a common problem for all nonlinear latent variable models.

2.2.1 Classical Estimation Methods

Recent development in statistical technology has made the estimation of nonlinear latent variable models possible with the increasing computing power. Among them, the Generalized Method of Moments (GMM) and Quasi-Maximum-Likelihood Estimation (QML) are two popular methods currently being applied. Other computationally-extensive procedures include simulated maximum likelihood(Daniellson 1994), efficient method of moments (Gallant, Hsieh and Tauchen 1995).

Method of Moments is a popular tool in econometrics. Applications of this method to the stochastic volatility model includes Chesney and Scott (1989), Melino and Turnbull (1990) and Anderson and Sorensen (1995). The main idea of the Method of Moments is to exploit the stationary and ergodic properties of the process which yield the convergence of sample moments to their unconditional expectations. It is useful in cases when distributional assumptions of the parameters are difficult if not impossible to make. For a fully specified parameter model like the stochastic volatility model, the Generalized Method of Moments method is expected to be inefficient.

Furthermore, there are a number of reasons that make the GMM approach an unattractive approach for the stochastic volatility model:

- GMM can only be used if the log-volatility process is stationary. When the mean reversion β is close to 1, which is common for most financial data series, GMM is expected to work poorly.
- Parameter estimates of GMM are not invariant. This is important if other

parameterizations are more interesting.

- GMM does not deliver an estimate of the volatility process h_t, either filtered or smoothed. Since the main use of the model is to forecast volatility, another estimation procedure has to be developed for that task.

Jacquier, Polson and Rossi provided the accuracy of the GMM estimates in their 1994 work and demonstrated its a less efficient estimation procedure for the stochastic volatility model.

Harvey, Ruiz and Shephard (1994) applied a Quasi-Maximum-Likelihood (QML) procedure to estimate the SV model. The basic model can be transformed into a linear state-space model by taking the logarithm of the squares of the observations.

$$log{y_t}^2 = h_t + log{\epsilon_t}^2, \tag{2.3}$$

$$h_{t+1} = \alpha + \beta h_t + \sigma_h u_t. \tag{2.4}$$

By using a normal approximation to the $log\chi^2(1)$ distribution, Harvey, Ruiz and Shephard approximated the model to a Gaussian state-space model and employed the standard Kalman Filtering technique to estimate the latent volatility process.

The attraction of QML is that it is very easy to implement and extends easily to more general models. And it provides filtered and smoothed estimates of the latent volatility process. But it is expected to have poor small sample properties since the $log\chi^2(1)$ distribution is poorly approximated by the normal distribution as shown in Figure 2.1. Jacquier, Polson and Rossi (1994) provided some empirical evidence on the performance of QML and concluded that as σ_h decreases, the QML procedure gave poor sampling properties.

We propose MCMC simulation techniques to conduct finite sample inference on the exact distribution of parameters of the Stochastic Volatility Model. Comparing

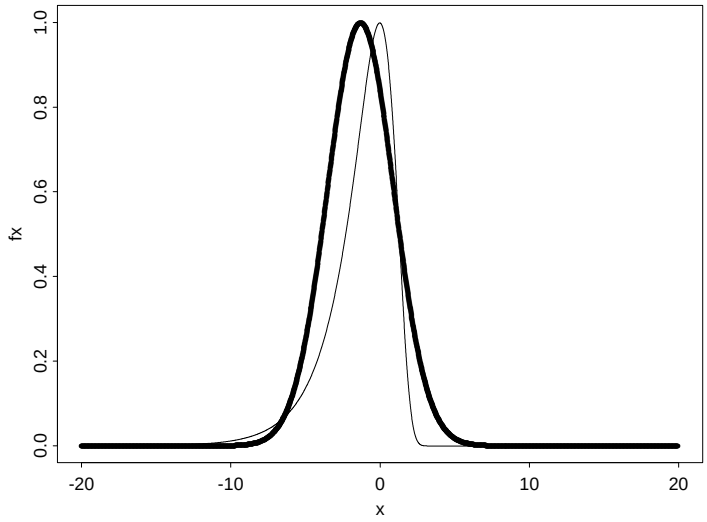

Figure 2.1: *The log$\chi^2(1)$ distribution and its normal approximation*

to the GMM and QML estimation procedures which rely on asymptotic approximations, Bayesian Markov Chain Monte Carlo (MCMC) estimation procedure offers a unified simulation approach which works on the actual likelihood distribution rather than asymptotic approximations for parameter estimation and smoothing. Jacquier, Polson and Rossi (1994) was among the first one who applied the MCMC technique to the Stochastic Volatility Model. They conducted extensive empirical studies on various data series and provided compelling evidence that the sampling performance of the MCMC sampler is much superior than that of the GMM and QML estimation procedure.

In this chapter, we will modify the MCMC sampler proposed by Jacquier, Polson and Rossi (1994) and propose a much faster MCMC sampler which uses recent sampling techniques developed for State-Space Models.

2.2.2 A Single-move Gibbs Sampler

In the Bayesian context, the Stochastic Volatility Model is treated as a hierarchical model. The latent volatility process is treated as a sequence of parameters in the model. This idea is termed data augmentation and was pioneered by Tanner and Wong (1987). Let $\omega = \{a, b, \sigma_h\}$ denote the space of model parameters, let h denote the vector of the volatility process, upon augmenting the volatility process $\{h\}$ to the parameter space $\{\omega, h\}$, the joint posterior distribution of ω, h can be written as the product of three conditional distributions. Namely, $p(\omega, h|Y) \propto p(Y|h)p(h|\omega)p(\omega)$. Upon assigning a prior distribution to ω, we can design a Gibbs Sampler which iteratively re-simulate the two conditional distributions:

- $p(h|Y, \omega)$,
- $p(\omega|h, Y)$,

and converges to the joint distribution of $p(\omega, h|Y)$. The marginal distribution $p(\omega|Y)$ can then be used to make inferences about model parameters. The marginal distribution $p(h|Y)$ provides the solution to the "smoothing" problem of inferring the latent volatility process as a natural by-product.

Now we study ways to sample from each conditional distribution.

Sampling from the second conditional distribution, $p(\omega|h, Y)$, is easy from standard Bayesian inference on linear models. Given the volatility process, the second equation in the model is a linear AR(1) regression model with unknown slope and intercept. We specify a conjugate prior for $\{\beta, v\}$ where $\beta = \{a, b\}$ and $v = \sigma_h^2$. Upon assigning a reference prior $p(\beta, v) \propto 1/v$, the posterior $p(\beta, v|h)$ can be easily derived from Bayesian linear models theory. In detail,

$$p(\beta, v|h) = N(\beta|\hat{\beta}, (X'X)^{-1}v)Ga(1/v|(n-2)/2, ns^2/2)$$

where n is the number of observations, $X_{(n-1)*2}$ is an $(n-1)$ dimensional matrix

$(1, h_t)_{n-1}$, $t = 1, ..n-1$, and ns^2 is the standard residual sum of square. Clearly, under the same framework, extension to higher order AR models and adding other regressors including seasonal dummy variables to either the volatility or the return process is straightforward using Bayesian linear model theory.

Sampling from the first conditional distribution, $p(h|Y, \omega)$, however, is not an easy work because of the log-normal structure. One way to sample the joint distribution as suggested in Jacquier, Polson and Rossi (1994) is to break the joint distribution $p(h|Y, \omega)$ into a set of univariate conditional distributions $p(h_t|h_{-t}, Y, \omega)$, where h_{-t} denotes the rest of the h vector other than h_t. Iterating through draws from each univariate conditional $p(h_t|h_{-t}, Y, \omega)$ distributions will converge to the joint distribution $p(h|Y, \omega)$. This is called a Single-Move Gibbs Sampler since sampling of the joint distribution is achieved by iterating through each univariate conditionals $p(h_t|h_{-t}, Y, \omega)$

Sampling of each marginal conditional distribution $p(h_t|h_{-t}, Y, \omega)$ is not easy because of the log-normal structure. A Metropolis algorithm is suggested by Jacquier, Polson and Rossi (1994). Because the Metropolis algorithm is going to be used T times for each univariate conditional distribution for each iteration of the Gibbs Sampler, and the Sampler will be iterating another N times, a highly effective proposal distribution for the Metropolis algorithm is needed. Notice that the univariate conditional distribution is a product of normal and log-normal distributions,

$$p(h_t|h_{t-1}, h_{t+1}, y_t, \omega) \sim p(y_t|h_t)p(h_t|h_{t-1})p(h_{t+1}|h_t).$$

This density can be very well approximated and dominated by a product of two inverse Gamma densities. Since the product of two inverse Gamma is still an inverse Gamma, an inverse Gamma distribution is chosen to be the Candidate Generating Distribution for the Metropolis algorithm. Experiments show a acceptance rate of $70\% - 80\%$ for the Metropolis algorithm which indicates the inverse Gamma density

is a good approximation to the conditional density.

Another possible way to sample $p(h_t|h_{-t}, Y, \omega)$ is to use accept/reject sampling algorithm as suggested by Geweke (1994). This can be achieved because that $p(h_t|h_{-t}, Y, \omega)$ is globally concave, consequently the general approach to acceptance sampling from distributions with log-concave distribution suggested by Wild and Gilks (1993) is applicable. This procedure is considerably faster than the Metropolis algorithm.

In summary, the Single-Move Gibbs Sampler iterates through the following steps:

1. Initialize h and ω.
2. Sample each h_t from $p(h_t|h_{-t}, Y, \omega)$.
3. Sample from $p(\omega|h)$.
4. Go back to 2.

It is easy to prove that the Single-Move Gibbs Sampler is ergodic and converges to the stationary distribution. But the actual rate of convergence has to be accessed by experimentation. We simulate data from the stochastic volatility model with known parameters and estimate the convergence rate of the sampler. After extensive experiments, we find that the sampler converges considerably fast when the autoregressive parameter b is small or the variance of the volatility process σ is large. But when the mean reversion parameter b is high, or σ is small, the sampler converges very slow.

To illustrate this, we simulate 1000 data from the model with a low b value and relatively large σ and study the convergence behavior of the sampler. We then simulate 1000 data with the same starting value but with a high b value and relatively small σ and compare the convergence of the sampler.

Figure 2.2 shows the simulation trajectory for b, the autocorrelation plot of the samples and the histogram of b for both series. It can be seen that the Sampler converges relatively faster for the series with $b = 0.5$ and $\sigma = 0.3$. The autocorrelation of the samples dies out after lag 400, and the chain converges to the stationary stage

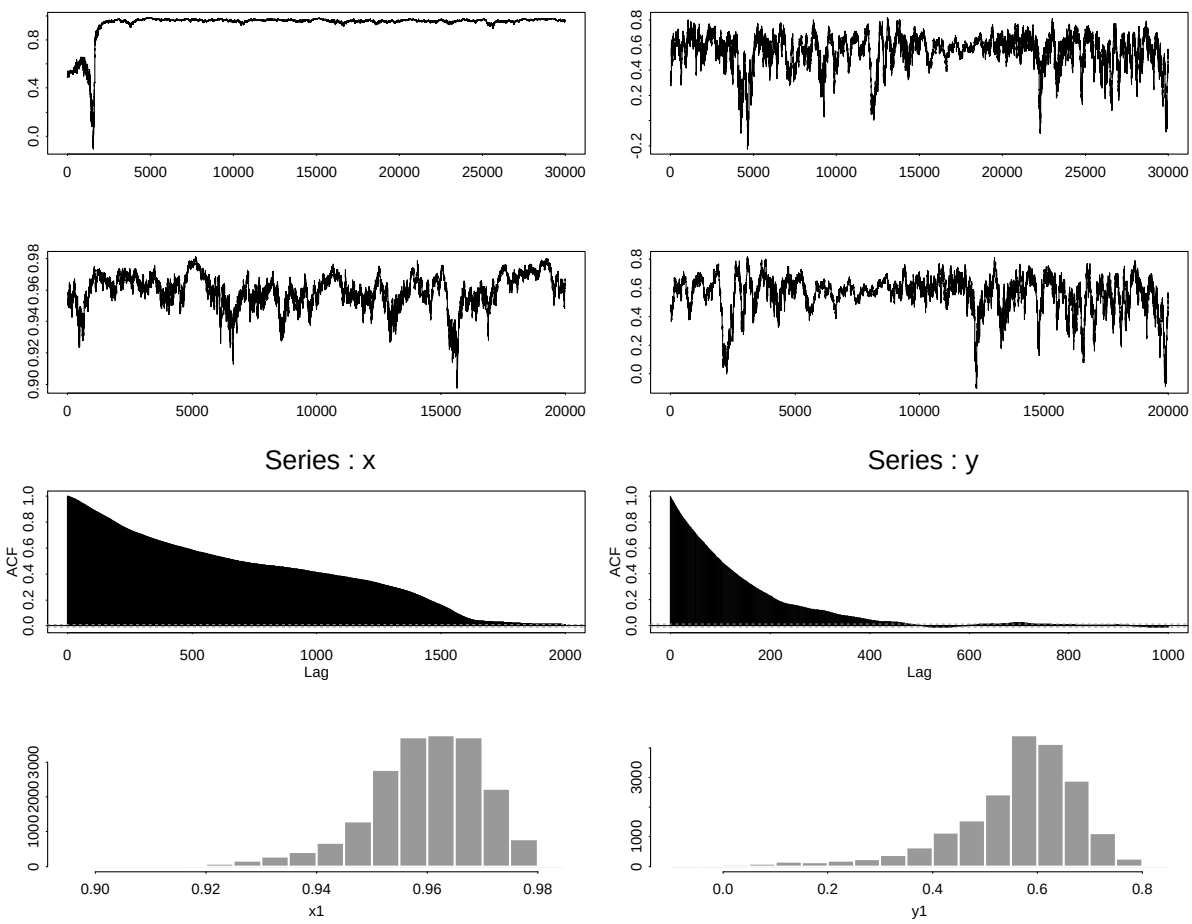

Figure 2.2: *Left: Trajectory plots, acf and histogram of b for a Single-move Gibbs Sampler with* $b = 0.95$, $\sigma = 0.1$. *Right: comparison with* $b = 0.5$, $\sigma = 0.3$.

after iterations 2000. For the series with $b = 0.95$ and $\sigma = 0.1$, the autocorrelation of the samples remains significant until lag 1500, and the chain converges to the stationary stage after iterations 5000.

This is because when b is high, or σ_h is very small, adjacent volatilities are highly correlated. This makes the mixing of the Gibbs Sampler in the parameter space very slow. When b is bigger than 0.9, significant autocorrelations in the samples of Gibbs Sampler remains even after lag 400. This unfortunate characteristic of the single-move Gibbs Sampler is common to all parameter-driven models, see Carter and Kohn (1994). If a component, such as h_t, changes slowly and persistently, the single-

move Gibbs Sampler will converge slowly. In the limit, when $h_t = h_{t-1}$, the sampler will not converge at all. Given that volatility in financial series tend to be highly correlated, this suggests that the Single-move Gibbs Sampler may be unreliable for real financial data.

2.2.3 A Multi-move Gibbs Sampler

A common solution to the slow mixing problem caused by highly correlated elements is to block the highly correlated elements and sample them together (see Smith and Roberts, 1993, and Liu, Wong and Kong, 1994). In the context of time series models, works on designing methods to sample blocks includes Carter and Kohn (1994) and Fruhwirth (1994). Instead of breaking the joint distribution $p(h|y,\omega)$ into several univariate conditionals $p(h_t|h_{_t},y,\omega)$ and sample from each univariate conditionals, we sample the joint distribution $p(h|y,\omega)$ directly.

To do this, the model

$$y_t = exp(h_t/2)\epsilon_t, \tag{2.5}$$

$$h_{t+1} = \alpha + \beta h_t + \sigma_h u_t, \tag{2.6}$$

is transformed into a linear state-space model.

$$log(y_t^2) = h_t + log(\epsilon_t^2), \tag{2.7}$$

$$h_{t+1} = \alpha + \beta h_t + \sigma_h u_t. \tag{2.8}$$

Here, $log(\epsilon_t^2)$ follows a $log\chi^2(1)$ distribution. After this transformation, Model (2.1) becomes a non-Gaussian dynamic linear model with h_t as the state vector. It is non-Gaussian because the error distribution for the observation process is a $log\chi^2(1)$

Table 2.1: Mean and Variance of the Mixture of Normals and the mixing probability

Mean	Variance	Mix P
-10.12999	5.79596	0.00730
-3.97281	2.61369	0.10556
-8.5668	5.17950	0.00002
2.77786	0.16735	0.04395
0.61942	0.64009	0.34001
1.79518	0.34023	0.24566
-1.08819	1.26261	0.25750

instead of a normal distribution. Yet, the $log\chi^2(1)$ distribution can be extremely well approximated by a seven component normal mixture following the idea of Carter and Kohn (1994). The degree of accuracy can be improved arbitrarily by increasing the number of mixtures. Following the work of Titterington, Smith and Makov (1985), we matched the first four moments of the seven component normal mixture density to that of the $log\chi^2(1)$ and require that the approximating densities lie within a small distance of the true density. The resulting mixing normals have means, variances and mixing probabilities displayed in Table 2.1.

Note that the mixture of normals are chosen independent of data. So this needs to be done only once.

After transforming the non-Gaussian state space model into a mixture of Gaussian State-Space model. The model becomes a standard Gaussian State-Space model at each time t. It is then possible to sample the entire joint distribution $p(h|y,\theta)$ at once using the standard State-Space simulation technique via Kalman filter. Now, in addition to the model parameter vector $\theta = \{\alpha, \beta, \sigma_h\}$ and volatility process $\{h\}$, we need to sample the mixing indicator series $K =_1, ...{}_N\}$, where each $k_t = 1, 2..7$.

The basic steps involves iterating through

1. Initialize θ, K.

2. Sample $p(K|Y,h,\theta)$.

3. Sample $p(h|Y, K, \theta)$.

4. Sample $p(\theta|h)$.

5. Go back to 2.

Sampling of $p(\theta|h)$ is the same as the Single-move Gibbs Sampler following standard Bayesian linear model theory.

To sample the indicator variables $K(t)$, we write down the joint likelihood of K given Y, h, θ.

$$p(K|Y, h) = p(k_N|Y^N, h) \prod_{t=1}^{N-1} p(K_t|Y^t, h_t, K_{t+1}).$$

Here Y^t consists of all y_i for $i \leq t$. Thus to generate K from $p(K|Y, h)$, we first generate k_N from $p(k_N|Y^N, h)$ and then for $t = N-1, ...1$, we generate k_t from $p(K_t|Y^t, h_t, K_{t+1})$. Because $p(k_N|Y^N, h)$ and $p(K_t|Y^t, h_t, K_{t+1})$ are discrete valued, we can generate from them easily, once we have calculated them.

To sample the state vector h all at once, we derive the joint likelihood function for h. For notational convenience, we omit the dependence on the indicator process K in the derivation of the following conditionals.

$$p(h|Y, \theta) \propto p(h_N|Y^N) \prod_{t=1}^{N-1} p(h_t|Y^t, h_{t+1})$$

Thus to generate $\{h\}$ from $p(h|Y, \theta)$, we first generate h_N from $p(h_N|Y^N)$ and then for $t = N-1, ...1$, we generate h_t from $p(h_t|Y^t, h_{t+1})$. Note that $p(h_N|Y^N)$ and each $p(h_t|Y^t, h_{t+1})$ are normal distributions, in order to generate the $\{h\}$ process, we only need to calculate $E(h_N|Y^N)$ and $var(h_N|Y^N)$, and $E(h_t|Y^t, h_{t+1})$ and $var(h_t|Y^t, h_{t+1})$ for $t = N-1, ...1$.

$E(h_N|Y^N)$ and $var(h_N|Y^N)$ can be obtained using Kalman Filter. To get $E(h_t|Y^t, h_{t+1})$

and $var(h_t|Y^t, h_{t+1})$, we treat the equation $h_{t+1} = a + bh_t + \mu_{t+1}$ as an additional observation on the state vector h_t and apply the Kalman Filter again. In other words, the sampling of the joint likelihood involves two stages. The first is forward filtering to calculate $E(h_N|Y^N)$ and $var(h_N|Y^N)$ and the second is backward smoothing to calculate $E(h_t|Y^t, h_{t+1})$ and $var(h_t|Y^t, h_{t+1})$. For details of derivation, see Carter and Kohn (1994).

One remaining issue is the specification of a prior for the initial state vector h_1. We choose a diffuse prior $h_1 \propto N(0, S_0)$ with a large S_0.

The convergence of the Multi-Move Gibbs Sampler is much faster even with high values of b and small values of σ_h. To compare with the Single-move Gibbs Sampler, we apply the Multi-Move Gibbs Sampler to the same data series generated in the previous section for $b = 0.95$ and $\sigma = 0.1$. Figure 2.3 shows the corresponding results. The sampler converges after 500 iterations. The autocorrelations in the samples of the Multi-Move Gibbs Sampler die out much faster than those in the Single-Move Gibbs Sampler. And the simulation time is also much shorter. The simulation time for $10,000$ iterations of the Multi-Move Gibbs Sampler is less than 10 minutes on an Alpha 3000 whereas that for the Single-move sampler is about 30 minutes. For $b = 0.95$ and $\sigma = 0.1$, the autocorrelations in the samples die out after lag 350 comparing to $1,500$ in the Single-move sampler.

2.3 Application to Exchange Rate Data

Volatility clustering, which is one of the most important motivations for the development of GARCH type models and Stochastic Volatility Models, is most obvious in stock/stock index return data and exchange rate data. We apply the model to analyze daily returns on the Deutschemark (DM)/US Dollar exchange rate from 1/3/81 to 5/31/96 and compare with existing research results on the same data set with

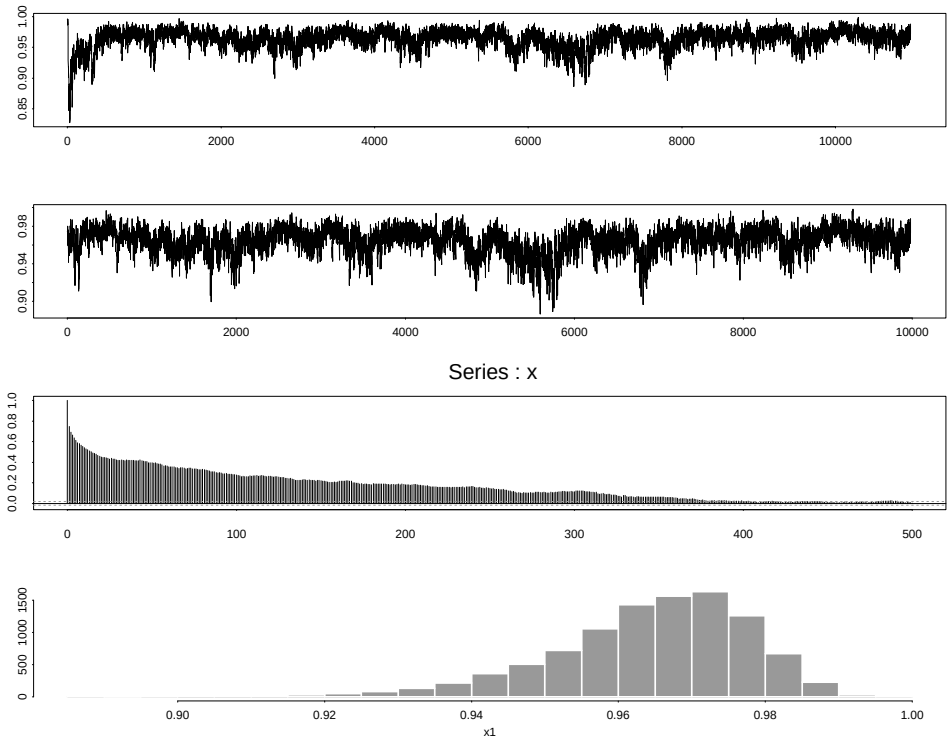

Figure 2.3: *Trajectory plot, acf and histogram of b for a Multi-move Gibbs Sampler with* $b = 0.95$, $\sigma = 0.1$

EGARCH model at J.P. Morgan. The DEM/USD data for a similar time period from a different data source was also studied by Jacquier, Polson and Rossi (JPR)(1994).

We choose exchange rate data because exchange rate typically exhibits high degree of volatility clustering. Another reason for using currency data over stock price series is that for major currencies, the leverage effect which the current model does not capture is less pronounced.

We use daily returns of the exchange rate because daily returns exhibits stronger degree of short-term volatility clustering than intra-day data and is less noisy. And the weekend effect is less important for currency data than for stock return data. Therefore, the need for using weekly data to avoid such minor problems is less important.

Both the Single-Move Gibbs Sampler and the Multi-Move Gibbs Sampler were

Table 2.2: Data: Parameter Estimates for Single-move and Multi-move Gibbs Sampler

	Single-move		*Multi-move*	
	Mean	Variance	Mean	Variance
$\beta \mid Y$	0.963	0.02	0.9627	0.02
$\sigma_h \mid Y$	0.168	0.04	0.168	0.04
$\alpha_h \mid Y$	-0.315	0.02	-0.317	0.02
Time	*115 minutes*		*28 minutes*	

applied to the data. We experimented with different starting values for h and ω. The results show that the effect of initial values disappear quickly after the first few thousand runs of the Gibbs Sampler. We run the Single-Move Gibbs Sampler for 110,000 iterations and discard the first 10,000 iterations. We then take a sample from every 200 samples to form a sample size of 500. This ensures that they are close to *iid* samples from the posterior distribution. These samples are then used to conduct inference on the parameters. For the Multi-move Gibbs Sampler, we run the chain for 80,000 iterations and discard the first 10,000 iterations. We then take a sample from every 100 samples to form a sample size of 800. Figure 2.4 shows the histogram of the posterior distribution of the parameters, the simulation trajectory and autocorrelation in the samples for both the Single-move and the Multi-move Gibbs Sampler. Table 2.2 shows the results of parameter estimates and simulation time. As can be seen, the autocorrelation in samples from the multi-move Gibbs Sampler is much smaller than that from the Single-move Gibbs Sampler and therefore requires much less iterations. The total simulation time for the Single-move sampler is about 2 hours for 1000 data sets whereas that for the Multi-move sampler is 30 minutes on a Sparc 20.

The results show the high values of the mean reversion parameter posterior mean of b is 0.96 and small variance for the volatility process, $\sigma^2 = 0.04$. These are consistent with JPR(1994)'s findings on the DM/Dollar data. JPR(1994) also provides an

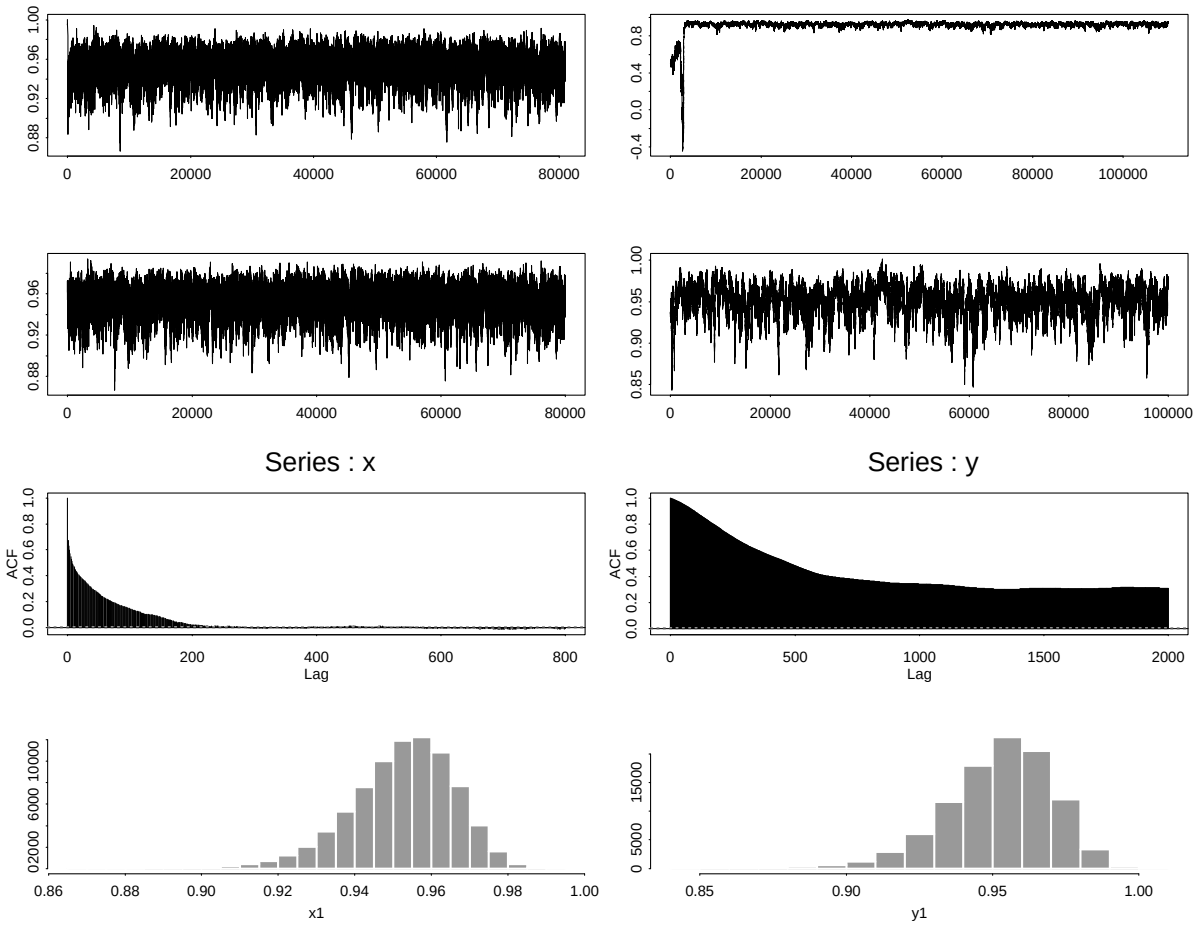

Figure 2.4: *Comparison of trajectory plot, acf and histogram of b for the DM/Dollar data using both the Single-move and Multi-move Gibbs Sampler*

extensive comparison on the accuracy of MCMC estimates and GMM and QML estimates. It is impossible to directly compare the parameter estimates of the Stochastic Volatility Model and EGARCH model since they have different parameterizations. But both models suggest high degree of volatility clustering in the DM/Dollar data.

2.4 Model Diagnostics

To evaluate the in-sample performance, we develop some diagnostic tools to check model adequacy for the Stochastic Volatility Model. Standard model diagnostic

procedures include residual analysis. The difficulty of residual analysis with every volatility forecasting model is that volatility is unobserved. This can be remedied by studying the properties of the squared returns. An additional difficulty with the residual analysis of Stochastic Volatility Models is the lack of analytic forms of the likelihood function. So simulation based diagnostic methods need to be developed.

Traditional residual analysis includes the check of model assumptions. In the Stochastic Volatility Model, model assumptions include normality of the return and volatility process and autocorrelation in the return residuals defined as $\epsilon_t = y_t/exp(h_t/2)$. In the Bayesian context, residual analysis are conducted under the distributional form. From the MCMC simulation of model estimation, we have the joint posterior distributions of the volatility process $p(h_0, ...h_T|Y)$, therefore, the joint posterior distributions of the residuals $p(e_0, ...e_T|Y)$ where $\epsilon_t = y_t/exp(h_t/2)$ are available. These can be used to calculate the distribution of autocorrelations of any lag, kurtosis, skewness, Box-Ljung and other statistics for normality and autocorrelation.

If the model captures the behavior of the data well, the residuals sequence $\epsilon_{t,j}$, $t = 1, ...T$ for each j should be independent. This can be checked by computing the autocorrelation of any lag for each residuals sequence $e_{t,j}$ $t = 1, ...T$ and then plotting the distribution of autocorrelation of all lags. The mean of the distribution of the autocorrelation of any lag can be calculated. A useful test statistics for checking the autocorrelations of a time series is the Box-Ljung test statistics. We also report this test statistics.

From the distribution of $p(\epsilon_0, ...\epsilon_T|Y)$, we can also calculate the distribution of the skewness, kurtosis among other test statistics to check the normality assumption.

Figure 2.6 shows the distribution of autocorrelation of lags 1 to 9 for the DEM/USD data. There is no significant autocorrelation in any lag. Figure 2.5 shows the the distribution of skewness and kurtosis statistics. The mean of the kurtosis statistics

is 3.38, a 95% confidence interval for the kurtosis statistics is $(2.54, 5.23)$. A 95% confidence interval for the skewness statistics is $(-0.81, 0.85)$. They indicate there is no excessive skewness or fat-tail behavior in the residuals. The QQ-plot in Figure 2.7 indicates the normality assumption is basically satisfied although there are some outliers.

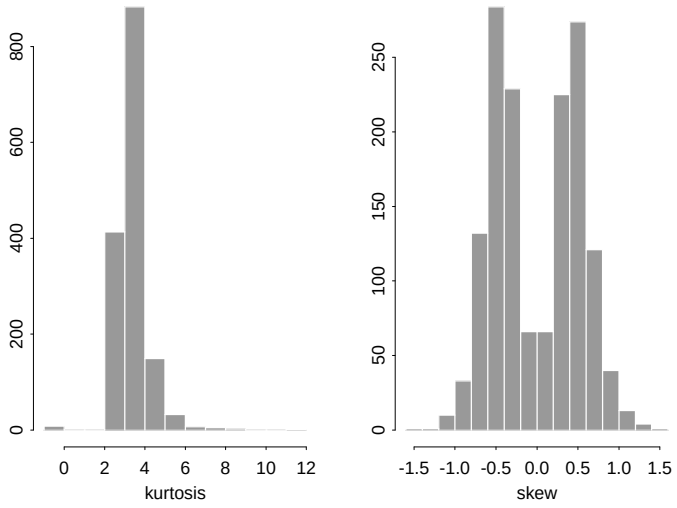

Figure 2.5: *Histogram of the skewness and kurtosis statistics for the DEM/USD Data*

A more ambitious in-sample diagnostic for a time series forecasting model is to evaluate the model's ability to capture the forecasting distribution of market variables. This is related to the Value at Risk (VaR) concept in risk management. The heart of market risk management is the forecast of the distribution of the relevant market variables. If an institution is estimating the Value at Risk in its trading and/or investment portfolio on a daily basis, it is implicitly forecasting the entire joint distribution of the market variables it is exposed to. Similarly, a forecast of a distribution is the central input into any asset allocation and/or hedging model. In this spirit, model diagnostics for any time series forecasting model should reveal the ability of the model to capture the distributional form of the variables forecast.

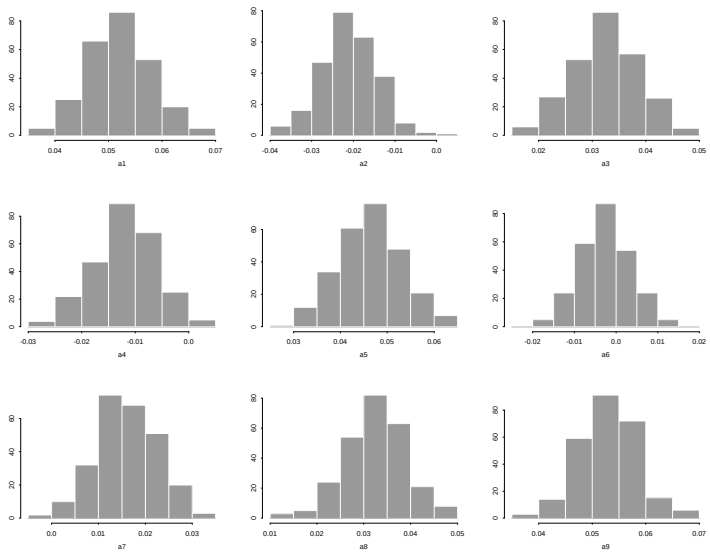

Figure 2.6: *Histogram of acf of the residuals of lag 1 to 9 for the DM/Dollar Data*

This can be checked by computing the percentiles of each new observation under the corresponding forecasting distribution. If the forecasting distribution captures the distribution of the market variable accurately, the percentiles should be independently and uniformly distributed. Independence measures how well the forecast model identifies the rapidly changing structure of the market; uniformity measures how well the forecasting model captures the shape of the distribution of the market variable over time. For a further discussion on this, see Crnkovic and Drachman (1996) from J.P. Morgan.

In the context of Stochastic Volatility Model, this requires the calculation of the predictive distribution $p(y_{t+1}|Y_t)$ where $Y_t = \{y_1, ..y_t\}$. This can be done using posterior samples of $\{h, \theta\}$ from the Gibbs Sampler. Note that

$$p(y_{t+1}|Y_t) = \int p(y_{t+1}|Y_t, h_{t+1}, \theta)p(h_{t+1}|Y_t, h_t, \theta)p(h_t|Y_t)p(\theta|Y_t)dh_{t+1}dh_t d\theta$$

So, $p(y_{t+1}|Y_t)$ can be sampled by the method of composition as follows. For each h_t^j, $j = 1, ...M$ from $p(h_t|Y_t)$ and each θ from $p(\theta|Y_t)$, we sample h_{t+1}^j from

$$h_{t+1}^{j} \sim N(\alpha + \beta h_t^j, \sigma_h)$$

Based on the M draws on h_{t+1}, we can estimate the probability that y_{t+1}^2 will exceed the observed y_{t+1}^{o2} by

$$Pr(y_{t+1}^2 \leq y_{t+1}^{o2}) = 1/M \sum_{j=1}^{M} Pr(y_{t+1}^2 \leq y_{t+1}^{o2} | h_{t+1}^j)$$

for each $t = 1, ..n$.

This approach can be extended to an arbitrary step ahead forecasting. Comparing to the traditional filtering framework where $p(h_{t+1}|Y_t, \omega)$ are calculated using an estimated ω, the Bayesian predictive distribution approach eliminates the parameter estimation error.

Once the simulated predictive distribution $p(y_{t+1}|Y_t)$ is available, we can compute the sequence of percentiles for the observed y_{t+1}^{obs}. This sequence should be approximately *iid* uniformly distributed. These variables can then be transformed into a normal distribution using the inverse of the normal distribution function. This sequence of standard *iid* normal variables can then be used to carry out Box-Ljung, normality and heteroscedasticity tests, among others.

The predictive density procedure was applied to the DEM/USD data set. Table 2.3 shows the results from the diagnostic checks on the model. These results are consistant with the Bayesian residual analysis and suggest no obvious failures in the way the model has been fitted. In particular, there is no excess kurtosis in the residuals which means the Stochastic Volatility Model is capable of generating excessive kurtosis the data exhibits. In comparison, the kurtosis generated by the EGARCH on the same data is not enough. (From personal communication with researchers at J.P. Morgan)

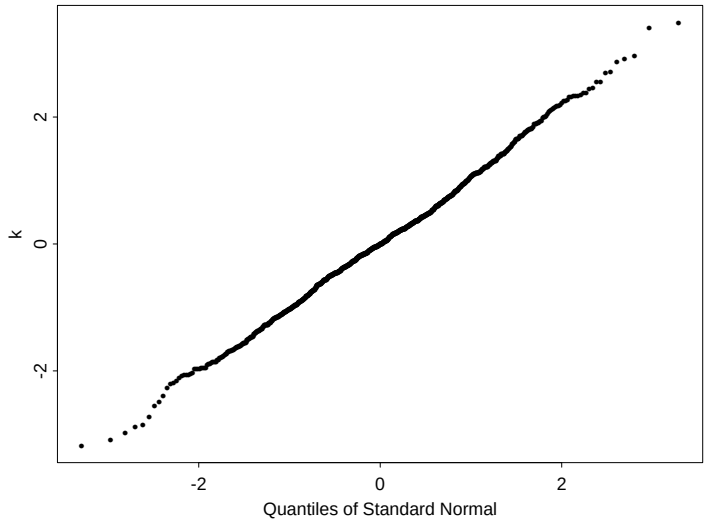

Figure 2.7: *QQ-plot of the sequence of normal variables from the predictive residual diagnostic of the DM/Dollar Data*

Table 2.3: Statistics from the predictive residual diagnostic

Skewness	Kurtosis	BL(30)	Normality
0.65	3.13	27.56	2.13

2.5 Conclusion

In this chapter, we proposed a Log-AR(1) Stochastic Volatility Model as a volatility forecasting model for financial data series. We developed two Bayesian Markov Chain Monte Carlo (MCMC) estimation procedures for the model and applied the model to some exchange rate data series. We developed both classical and Bayesian model diagnostic tools to check in-sample model adequacy for the real data. This represents a complete study for the development of a financial forecasting model. Our study indicates that the Log-AR(1) Stochastic Volatility Model is a promising volatility forecasting model comparing with existing volatility forecasting models. Comparing to the EGARCH model, the Log-AR(1) Stochastic Volatility Model is capable of generating excessive kurtosis exhibit in the financial data. In the following chapters,

we will discuss extensions to the model and procedures for out-of-sample forecasting using the model and the problem of option pricing using the model.

Chapter 3

Model Extension

In this chapter, we propose three extensions to the simple Log-AR(1) Stochastic Volatility Model. In the first section, we implement a Log-AR(p) Stochastic Volatility Model and apply it to some real exchange rate data. In the second section, we discuss the leverage effect of volatility and the work of Jacquier, Polson and Rossi (1995) on the correlated Stochastic Volatility Model and Breidt (1996) on a threshhold Stochastic Volatility Model. In the third section, we propose a new model which combines option pricing model and stochastic volatility model under one model framework. Unlike the previous model specifications which solely model the behavior of historical volatility, this model combines historical and implied volatility and provides a unified approach to forecast volatility. This new model can be used for both forecasting and testing of the hypothesis of the existence of stochastic volatility.

3.1 A Log-AR(p) Stochastic Volatility Model

Volatility clustering is one motivation of applying Autoregressive Stochastic Volatility Models to forecast volatility. Therefore, exploring autocorrelations in the volatility process is the central role in model building. Although no autocorrelation was found in the residual of Log-AR(1) Stochastic Volatility Model in DM/Dollar data, it is still worthwhile to explore a Log-AR(p) Stochastic Volatility Model as an alternative way to study further autocorrelation structure in the data. The significance of the higher order AR(p) parameters will suggest the existence of higher order autocorrelations in

the volatility process. As stated in the last chapter, it is conceptually easy to extend the model fitting procedure to higher order AR(p) process for the volatility. It is worth while implementing the procedure and examine the fit of a higher order Log-AR(p) Stochastic Volatility Model to the data. We start with a Log-AR(2) Stochastic Volatility Model:

$$r_t = exp(h_t/2)\epsilon_t, \tag{3.1}$$

$$h_{t+1} = \alpha + \beta_1 h_t + \beta_2 h_{t-1} + \sigma_h \mu_t. \tag{3.2}$$

In this model, the log-volatility process is an AR(2) process. The significance of the AR(2) parameters will suggest the existence/nonexistence of the second order autocorrelation in the volatility.

Both the Single-Move Gibbs Sampler and the Multi-Move Gibbs Sampler can be extended easily to the Log-AR(2) Stochastic Volatility Model. We will discuss the details of applying the Multi-Move Gibbs Sampler here.

After transforming the model into a non-Gaussian State-Space model, we use a mixture of normal distributions to approximate the $log\chi^2(1)$ distribution as before and transform the model into a Gaussian State-Space model at each time t.

Re-write the model as a standard State-Space Model:

$$y_t = F\theta_t + v_t, \tag{3.3}$$

$$\theta_{t+1} = G\theta_t + \omega_t, \tag{3.4}$$

where $y_t = log(r_t^2)$, $\theta_t = (h_t, h_{t-1}, 1)'$, $F_{1*3} = (1,0,0)$, $G_{3*3} = \begin{pmatrix} \beta_1 & \beta_2 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $v_t \sim N(\mu_i, \sigma_i^2)$ at each time t with a pre-specified (μ_i, σ_i^2), $\omega_t = (\omega_t, 0, 0)'$ has a trivariate

normal distribution with zero mean and variance-covariance matrix $\begin{pmatrix} \sigma_h^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

The two error sequences v_t and ω_t are mutually independent.

Before we apply the multivariate version of the Multi-move Gibbs Sampler, we study the asymptotic distribution for $p(\alpha, \beta_1, \beta_2|Y)$. To do so, we transform the non-linear Log-AR(2) model to a linear State-Space model,

$$log(y_t^2) = h_t + \epsilon_t \tag{3.5}$$

$$h_{t+1} = \alpha + \beta_1 h_t + \beta_2 * h_{t-2} + \mu_t \tag{3.6}$$

Here $\epsilon_t \propto log\chi^2(1)$, the variance v_1 of the $log\chi^2(1)$ is 4.6 and the variance $\mu_t \propto N(0, v_2)$. In most financial data series, v_2 is less than 0.3. So the variance of the first equation is at least 15 times bigger than the variance of μ_t in the second equation.

Plugging $h(t)$ into the first equation, we get

$$y_t = \alpha + \beta_1 h_{t-1} + \beta_2 h_{t-2} + \mu_t + \epsilon_t$$

This is not exactly true since y_t should also depend on μ_{t-1}. But the focus here is to study the variance scale of the marginal posterior distribution for $p(\alpha, \beta_1, \beta_2|Y)$. Since the variance of μ_t is much smaller than that of ϵ_t, $v2$ is negligible.

Consider this as a regression function in matrix form and assuming the $\{h\}$ sequence is known,

then $p(\alpha, \beta_1, \beta_2|Y) \propto N((X'X)^{-1}X'Y, (X'X)^{-1}v1)$

where $X = \begin{pmatrix} 1 & h_t & h_{t-1} \\ \vdots & \vdots & \vdots \end{pmatrix}$ is the $(N-2)*3$ matrix of $\{h\}$.

To estimate the scale of the variance-covariance matrix, we need to know the scale

of $(X'X)^{-1}$. $(X'X) = \begin{pmatrix} n-2 & \rho_{12} & \rho_{13} \\ \rho_{12} & v_2 & \rho_{23} \\ \rho_{13} & \rho_{23} & v_3 \end{pmatrix}$

where $\rho_{12} = \sum_{t=2}^{n-2} h_t^2$, $\rho_{13} = \sum_{t=2}^{n-2} h_t h_{t-1}$, $v_2 = \sum_{t=2}^{n-2} h_t^2$, $\rho_{23} = \sum_{t=2}^{n-2} h_t h_{t-1}$ and $v_3 = \sum_{t=1}^{n-1} h_{t-1}^2$.

We simulated 1000 data using $a = -0.6$, $b = 0.8$, $v2 = 0.04$, the variance-covariance matrix for the data is: $\begin{pmatrix} 0.55 & 0.0238 & 0.0231 \\ 0.0231 & 0.0233 & -0.0213 \\ 0.0231 & -0.0213 & 0.0233 \end{pmatrix}$. Increasing the number of observations decreases the variance. For $t = 10000$, the variance-covariance matrix for the data is:

$$\begin{pmatrix} 0.0399 & 0.00167 & 0.00166 \\ 0.00167 & 0.00241 & -0.00227 \\ 0.00167 & -0.00227 & 0.00242 \end{pmatrix}$$

Experiments with different parameter values and different number of observations show the scale of the variance-covariance matrix stays in that range. The smallest standard deviation for the marginal distribution of α , β_1 and β_2 is at the scale of 0.1. This is assuming known $\{h\}$. With unknown $\{h\}$, the standard deviations for the marginal distribution of α,β_1 and β_2 will be much bigger. This indicates the variances of the marginal distribution for α,β_1 and β_2 are very big.

These studies suggest that the joint posterior distribution $p(a, b|Y)$ has large variances comparing to the $AR(1)$ model. This is because the second autoregression coefficient introduces much more uncertainty in the parameter estimates.

We now apply the Multi-move Gibbs Sampler to some simulated data from the model. We simulate 1000 data from the model with $\beta_1 = 0.8$, $\beta_2 = 0.1$ and $\sigma_h = 0.2$. We choose a uniform distribution over a constrained region $\beta_1 + \beta_2 < 1$, $\beta_2 - \beta_1 < 1$, and $-1 < \beta_2 < 1$ to ensure stationarity.

Figure 3.1 shows the simulation trajectory and histogram of parameters. The autocorrelation in the samples is big until lag 2000 for the Multi-move Gibbs Sampler.

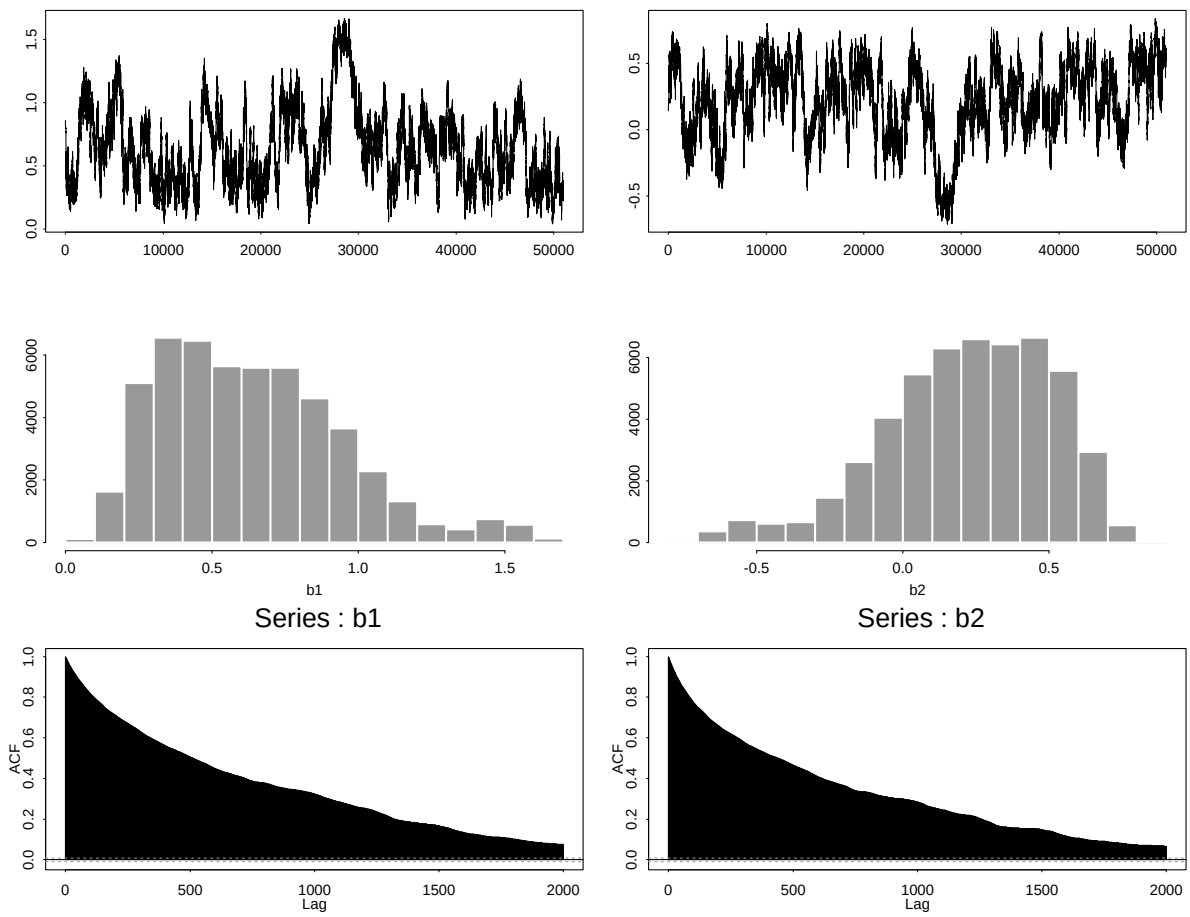

Figure 3.1: *Simulation trajactory, histogram and acf of* b_1 *and* b_2 *for the Log-AR(2) model with* $b_1 = 0.65$, $b_2 = 0.2$

This is much bigger than that in the Log-AR(1) model. The 95% confidence interval for β_1 is $(0.185, 1.43)$ the 95% confidence interval for β_2 is $(-0.5, 0.64)$. The mean for β_1 is 0.64, the mean for β_2 is 0.22. It can be seen that the uncertainty in the β_1 and β_2 is very big.

We then apply the Log-AR(2) model to the DM/Dollar data. The results are in Figure 3.2. The autocorrelation in the samples is similar to that using simulated data. The 95% confidence interval for β_1 is $(0.08, 0.97)$, the 95% confidence interval for β_2 is $(-0.01, 0.86)$. The mean for β_1 is 0.28, the mean for β_2 is 0.66. It is not intuitive that the lag 1 autocorrelation parameter is bigger than the lag 2 autocorrelation

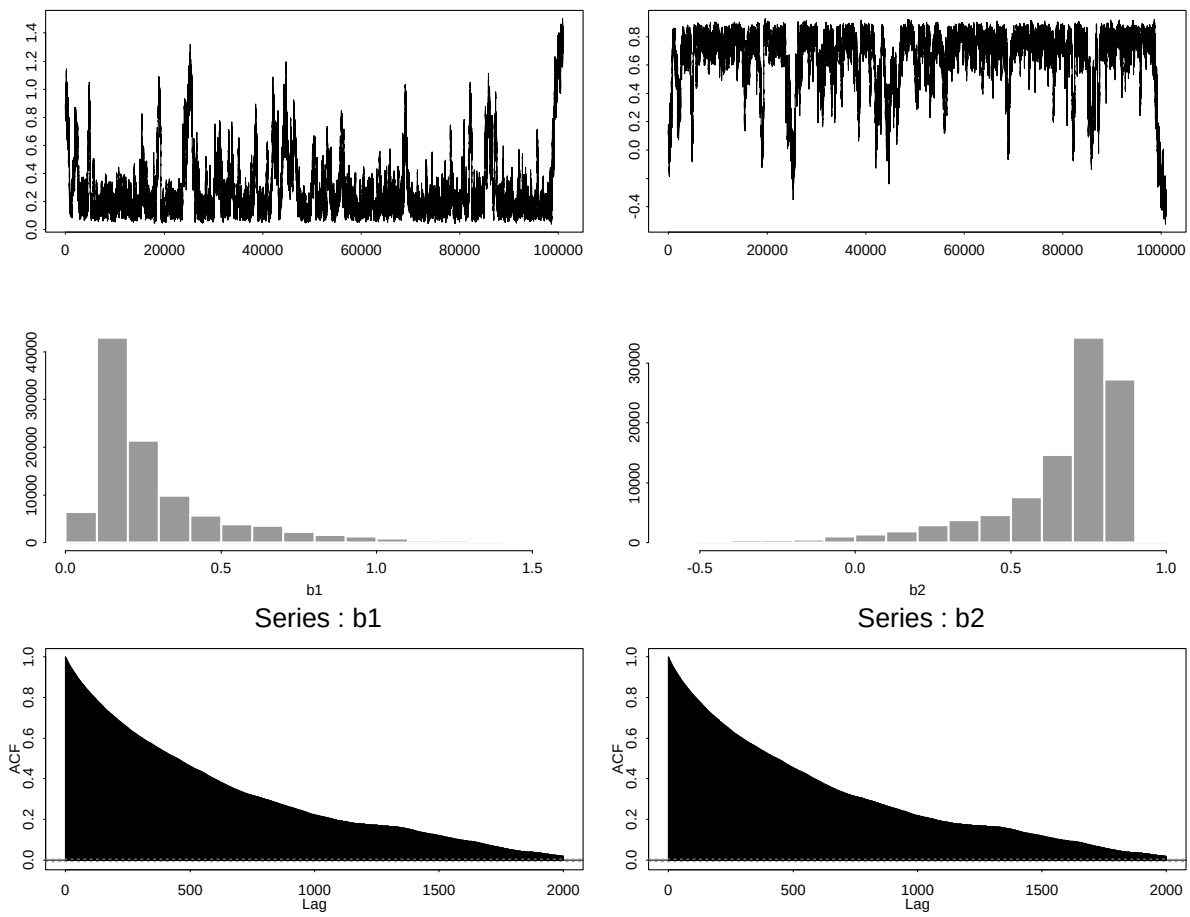

Figure 3.2: *Simulation trajectory, histogram and acf of* b_1 *and* b_2 *for DM/Dollar data with the log-AR(2) model*

parameter.

3.2 Correlated Log-AR(1) Stochastic Volatility Model

A more important extension to the simple Log-AR(1) Stochastic Volatility Model is to allow the correlation between the two innovations. This will create the so-called "leverage effect" which allows the sign and the value of the return to affect the value of the volatility. The intuition behind is that market is more volatile in a bear market

than in a bull market. The second extension to the model is to allow the correlation between the two error processes in the return and volatility process. This allows the asymmetric effect of return on volatility. Since the leverage effect is one of the three most important stylized effects of volatility found in financial data series, we discuss the work of Jacquier, Polson and Rossi (1995) on the correlated Stochastic Volatility Model and Breidt (1996) on a threshold Stochastic Volatility Model.

3.2.1 A Correlated Stochastic Volatility Model

Empirical study has shown the "leverage effect" is common in stock returns and in currency returns of emerging markets. In EGARCH model, the "leverage effect" is modeled by adding the return term into the equation for volatility. The significance of the parameter of the return will indicate the effect of the return on the value of the volatility. The Correlated Stochastic Volatility Model treats the "leverage effect" in a more elegant way by allowing the error distribution of the two processes to be correlated.

The model is:

$$y_t = exp(h_t/2)\epsilon_t, \tag{3.7}$$

$$h_{t+1} = \alpha + \beta h_t + \sigma_h \mu_t. \tag{3.8}$$

Now ϵ_t and μ_t have a bivariate normal distribution with variance-covariance matrix $\binom{1\ \rho}{\rho\ 1}$ where ρ denotes the covariance.

Estimation of this model now requires simulation from the joint posterior distribution $p(h, \omega, \rho|Y)$. Two difficulties arise. First, one needs to specify an appropriate joint prior distribution for the correlation ρ and volatility of volatility σ_h. Second, the simulation algorithm needs to be modified to accommodate the covariance ρ.

Jacquier, Polson and Rossi (1995) discussed ways to assign a joint prior for ρ and σ_h and developed a Single-Move Gibbs Sampler to estimate the correlated model. The main difficulty in specifying a joint prior is that the top left element in the variance-covariance matrix is equal to one. The standard conjugate inverse Wishart family prior can not be used because it is impossible to model beliefs where some elements of the matrix are known whereas others are not. To solve this problem, the model is re-parameterized as $\{\alpha, \beta, \Sigma\}$, where $\Sigma = \begin{pmatrix} 1 & \rho\sigma_h \\ \rho\sigma_h & \sigma_h^2 \end{pmatrix}$. Here, ρ denotes the correlation. The joint distribution of the data and the volatilities are

$$p(Y, h|\alpha, \beta, \Sigma) = \prod_{t=1}^{N} h_t^{-3/2} p(h_t^{-1/2} y_t, log h_t | h_{t-1}, \alpha, \beta, \Sigma)$$

Further rewrite the joint distribution in terms of residuals $(r_t, \mu_t)'$ where $r_t = h_t^{-1/2} y_t$ and $\mu_t = log h_t - \alpha - \beta log h_{t-1}$, for $t = 1, .., N$. The joint distribution of the data and the volatilities is now

$$p(Y, h|A, \Sigma) = \prod_{t=1}^{N} h_t^{-3/2} |\Sigma|^{-3/2} exp(-1/2 tr(\Sigma^{-1} A))$$

where $A = \sum_t r_t r_t'$ is the residual matrix.

Further rewrite the variance-covariance matrix with an hierarchical structure to the probability distribution. Let $\psi = \rho\sigma_h$ and $\Omega = \sigma_h^2(1 - \rho^2)$ with inverse transformation $\sigma_h^2 = \Omega + \psi^2$ and $\rho = \psi/\sqrt{\Omega} + \psi^2$. The joint prior can then be specified in a conditional fashion as $p(\Sigma) = p(\psi|\Omega)p(\Omega)$. This reparameterization can be viewed as reformulating the covariance matrix of r_t and μ_t in terms of the linear regression of μ_t on r_t, The new parameters ψ and Ω are the slope coefficient and variance of the error distribution. A natural joint prior distribution for the regression coefficients is a normal-gamma distribution as used in previous chapters.

A Single-Move Gibbs Sampler can be constructed to sample the joint posterior distribution $p(h, \alpha, \beta, \Sigma|Y)$ by iterating through the following three conditionals:

- $p(h|Y, \psi, \Omega, \alpha, \beta)$
- $p(\psi, \Omega|h, Y, \alpha, \beta)$
- $p(\alpha, \beta|h, Y, \psi, \Omega)$

For details of sampling $p(\psi, \Omega|h, Y, \alpha, \beta)$ and $p(\alpha, \beta|h, Y, \psi, \Omega)$, see Jacquier, Polson and Rossi (1995).

Unfortunately, the Multi-Move Gibbs Sampler can not be applied to estimate the correlated Stochastic Volatility Model because the log transformation of the return equation destroys the variance-covariance structure.

3.2.2 A Threshold Stochastic Volatility Model

An alternative specification for the leverage effect is the Threshold Log-AR(1) Stochastic Volatility Model dicussed by Breidt (1996). The Threshold Log-AR(1) Stochastic Volatility Model allows the log volatility process switches between two AR(1) processes according to the sign of the previous return. The model is:

$$y_t = exp(h_t/2)\epsilon_t, \tag{3.9}$$

$$\theta_{t+1} = \begin{cases} \alpha_1 + \beta_1\theta_t + \sigma_1\mu_{1t} & , \quad y_t \leq 0, \\ \alpha_2 + \beta_2\theta_t + \sigma_2\mu_{2t} & , \quad y_t > 0. \end{cases} \tag{3.10}$$

The volatility process is modeled as a threshold process with threshold variable y_t. μ_{1t} and μ_{2t} are *iid* normally distributed errors. ϵ_t and μ_t^j $(j = 1, 2)$ are independent.

One attraction of the Threshold Stochastic Volatility Model is the ease of specification of a prior distribution for the parameters. In addition, it allows additional dynamics of the volatility process by assigning different mean-reversion and drift parameters for positive and negative returns.

The main difficulty in the estimation of the correlated Stochastic Volatility Model is the specification of a prior distribution for the correlation parameter ρ. The Threshold Stochastic Volatility Model avoids this problem. The parameters α_1, α_2, β_1, β_2 and σ_1 and σ_2 can be considered as independent. A multivariate normal distribution with specified mean vector and diagonal variance-covariance matrix can be used as the joint prior distribution for α_1, α_2 and β_1, β_2. The prior distribution for σ_1^2 and σ_2^2 are independent inverse gamma distributions with pre-specified positive parameters.

Breidt (1996) proposed a Single-Move Gibbs Sampler for the Threshold Stochastic Volatility Model. Let $\theta = \{\alpha_1, \beta_1, \alpha_2, \beta_2\}$, $\sigma = \{\sigma_1, \sigma_2\}$, the sampler iterates through

- $p(\theta|Y, h, \sigma)$
- $p(\sigma|Y, h, \theta)$
- $p(h|Y, \theta, \sigma)$

To sample the h vector, the joint posterior $p(h|Y, \theta, \sigma)$ is broken into N univariate conditional posterior $p(h_t|h_{t-1}, h_{t+1}, Y, \theta, \sigma)$. Following Breidt (1996),

$$\begin{aligned} p(h_t|h_{t-1}, h_{t+1}, Y, \theta, \sigma) &\propto p(h_{t+1}|h_t, y_t, \theta, \sigma)p(y_t|h_t)p(h_t|h_{t-1}, y_t, \theta, \sigma) \\ &\propto k_t p^*(h_t), \end{aligned}$$

where k_t contains all factors not depending on h_t, so that

$$\begin{aligned} log p^*(h_t) &= -h_t/2 - (h_t - \mu_t)^2/(2\tau_t^2) - y_t^2 e^{-h_t}/2, \\ &\leq -h_t/2 - (h_t - \mu_t)^2/(2\tau_t^2) - y_t^2 e^{-h_t}/2\{1 - (h_t - \mu_t)\}, \\ &= -(h_t - \mu_t)^2/(2\tau_t^2) + (\mu_t^2 - \mu_t^2)/(2\tau_t^2) - k_t p(h_t) y_t^2 e^{-h_t}/2(1 + \mu_t), \end{aligned}$$

with $\tau_t^2 = (\sigma^{-2t} + \beta_{t+1}^2 \sigma_{t+1}^{-2})^{-1}$, $\mu_t = [\alpha_t + \beta_t h_{t-1}/\sigma_t^2 + (h_{t+1} - \alpha_{t+1})\beta_{t+1}/\sigma_{t+1}^2]\tau_t^2$ and $\mu_t^* = \mu_t + \tau_t^2/2(y_t^2 e^{-\mu_t} - 1)$.

This suggests a rejection algorithm to draw from each $p(h_t|h_{t-1}, h_{t+1}, Y, \theta, \sigma)$.

Breidt (1996) applied the Threshold Stochastic Volatility Model to the Center for Research in Security Prices (CRSP) data used by Nelson (1991) in his original EGARCH paper. He found that there is some evidence in the CRSP data that volatility following a positive return differs from volatility following a negative return, not only in level but also in dynamic behavior. There is some evidence that $\alpha_1 > \alpha_2$ and $\beta_1 > \beta_2$. This suggests that the mean reversion and drift parameters for the volatility process could be different for positive and negative returns. This difference is not captured by the constant parameter Correlated Stochastic Volatility Model.

3.3 Combining Historical and Implied Volatility in one Model

Up to this point, we are modeling historical volatility. We believe future volatility can be predicted using patterns from historical volatility. In the market, historical volatility is not the only source of estimate of volatility. The "implied volatility", computed from current option prices in the market, is another important source for volatility estimation. In we consider constant interest rate, volatility is the only unknown parameter in the option pricing formula. Therefore, after observing market option prices, a volatility estimate can be computed using the option pricing formula.

The implied volatility has two distinct features from historical volatility. First, it contains information of expectation towards the future since option price is the expectation of future price movements and volatility of the underlying asset. This makes the information from implied volatility very different from that of historical volatility. There is some empirical evidence that future expectation causes a different

behavior of "implied volatility" from historical volatility. For example, in 1995, a few months before the Canadian election, implied volatility from the Canadian dollar option increased significantly following the news of possible independence of Quebec whereas the historical volatility of the currency itself did not change much.

Second, the calculation of "implied volatility" is model dependent. Usually, the term "implied volatility" means "Black-Sholes implied volatility" which indicates the volatility is calculated from the Black-Sholes option pricing formula. But it is well known that the assumptions of the benchmark Black-Sholes option pricing formula is violated in the market. In fact, Black-Sholes assumes the volatility is constant. Option prices observed from the market could be derived from different option pricing formulas which is unknown to the public. Therefore, studies of the implied volatility using one option pricing formula are model dependent.

The debate over which volatility is a better estimate of the market expectation of the variance has existed long and has received mixed support from empirical evidence. The time series study of Lamoureux and Lastrapes (1993) considers options on non-dividend paying stocks and compared the forecasting performance of GARCH, implied volatility and historical volatility estimates and found that implied volatility forecasts, though biased, outperform the others. In sharp contrast, Canina and Figlewski (1993) studied $S\&P500$ index call options for which there is an extremely active market. They found that implied volatilities were virtually useless in forecasting future realized volatilities of the $S\&P500$ index. In a different setting using weekly sampling intervals for $S\&P100$ option contracts and a different sample, Day and Lewis (1992) not only found that implied volatilities had a predictive content but also were unbiased. Studies examine options on foreign currencies, such as Jorion (1995) also found that implied volatilities outperformed historical volatilities to predict future realizations. As to my best knowledge, no one has yet, combined historical

and implied volatility in a unified approach to forecast volatility and price options.

In this section, we propose a method which, for the first time, combines historical volatility and implied volatility under one model framework. We test the hypothesis of the existance of stochastic volatility and we investigate the possibility of a more accurate prediction of the future volatility and therefore a more accurate forecast of option price using the combined information from both the historical return and current option price.

We start with a simple stochastic volatility model developed by Hull and White (1987) to illustrate our idea and discuss extensions to more complicated dynamics of other stochastic volatility models.

3.3.1 The Model

We follow the stochastic diffusion model developed by Hull and White (1987) in this section to illustrate our idea and discuss ways to extend this idea to the Log-AR(1) Stochastic Volatility Model and other more complicated models in the following sections.

Hull and White (1987) proposed an option pricing model with stochastic volatility which assumes that the volatility itself is a state variable following a diffusion process. The model represents the class of stochastic volatility option pricing models, including those of Scott (1987), Wiggins (1987) and Johnson and Shanno (1987), that assumes volatility risk doesn't affect the option price. The continuous time process for the underlying asset is:

$$dS = \phi S dt + \sqrt{V} S d\omega, \tag{3.11}$$

$$dV = \mu V dt + \xi V dz, \tag{3.12}$$

where S is the asset price, V is the volatility of the return of the asset. Here ϕ is a parameter that may depend on S, σ and t, μ and ξ may depend on V and t, but not on S, and $d\omega$ and dz are two Brownian motions which have an instantaneous correlation ρ. Under the assumption that volatility risk is not priced, $\rho = 0$.

Under this setting for the return and volatility, Hull and White (1987) derived a formula for a call option as

$O_t = \int BS(\bar{V}_t) h(\bar{V}_t|V_t) d\bar{V}_t = E[BS(\bar{V}_t|V_t)]$

where $\bar{V}_t = \frac{1}{T-t} \int_t^T V_i di$.

The term $h(\bar{V}_t|V_t)$ in (3) is the density of $\bar{V}_t$ conditional on the current volatility V_t, T is the expiration date of the option, V_t is the volatility at time t and $BS(.)$ is the Black-Sholes pricing formula. Thus, the Hull-White price of the option is the mean Black-Sholes prices, evaluated over the conditional distribution of average variance $\bar{V}_t$. Under this framework, the "Black-Sholes implied volatility" can be interpreted as the average implied volatility.

Notice that the Hull-White option pricing formula is a function of the current volatility v_t, not the average volatility $\bar{V}_t$. Given v_t, the option price is completely determined.

Under this framework, we propose a volatility forecasting model which combines historical volatility and implied volatility in an unified way. We then discuss how to apply Markov Chain Monte Carlo simulation to the model and produce volatility estimates conditioning on both historical volatility and implied volatility.

A discrete approximation to the above continuous model is:

$$\begin{aligned} ln(S_{t+1}/S_t) &= \alpha - 1/2 V_t + \sqrt{V_t}\mu_t, \\ lnV_{t+1} &= \beta - \sigma_\epsilon^2/2 + lnV_t + \sigma_\epsilon \epsilon_t. \end{aligned}$$

Adding the option price formula into the model, we have a series of three processes regarding the return, option price and volatility:

$$
\begin{aligned}
ln(S_{t+1}/S_t) &= \alpha - 1/2V_t + \sqrt{V_t}\mu_t, && (3.13)\\
lnV_{t+1} &= \beta - \sigma_\epsilon^2/2 + lnV_t + \sigma_\epsilon\epsilon_t, && (3.14)\\
O_t^1 &= f(S_t, V_t) + \sigma_\omega\omega_t, && (3.15)
\end{aligned}
$$

where

$$
f(S_t, V_t) = \int BS(\bar{V}_t)h(\bar{V}_t|V_t)d\bar{V}_t = E[BS(\bar{V}_t|V_t])]. \quad (3.16)
$$

Here, μ_t, ϵ_t, ω_t are independent normally distributed, ω_t can be interpreted as market friction, bid-ask spread and other noises in the option prices, and O_t^1 indicates one option price at time t. We can add different options on the same underlying variable with different time to maturity to the model.

Notice that V_t appears in both the return process and option pricing formula, so that V_t is the combination of both historical volatility and implied volatility. Both the observed return and option prices contribute to the estimation of V_t and forecast of future option prices.

As stated earlier, one limitation of using implied volatility for volatility estimation is that implied volatility is model dependent. Empirical evidence shows that this is especially true for the commonly used "Black-Sholes Implied Volatility" derived from the Black-Sholes pricing formula. The commonly observed "smile effect" of volatility suggests the "Black-Sholes Implied Volatility" usually over-estimates volatility. However, the problem of model dependence is less severe in our model since the option prices are derived using the process for the return and volatility specified within the

model. If the underlying process for the return and volatility is a good approximation to what's observed in the market, forecasts of future option prices using this model will be more accurate than forecasts using historical or implied volatility alone.

Next, we develop a sampling scheme for the model using the Markov Chain Monte Carlo simulation framework. As will be seen later, MCMC sampling allows the estimation of volatility conditioning on both return and option prices in a unified approach.

3.3.2 MCMC for the Model

Again, estimation is the major difficulty of the study of this model. We now develop a MCMC sampling algorithm for the model using a combination of Gibbs Sampler and Metropolis algorithm. Consider model (3.14) and let $Y = (y_1, y_2, ...y_n)$ where $y_t = ln(S_{t+1} - S_t)$, $V = (v_1, v_2, ...v_n)$, $O = (O_1, O_2, .., O_n)$, and $\theta = \{\alpha, \beta, \sigma_\epsilon, \sigma_\omega\}$.

In the Bayesian context, the model setup leads to a hierarchical structure of conditional distributions. They are,

- $p(O|Y, V, \theta)$
- $p(Y|V, \theta)$
- $p(V|\theta)$
- $p(\theta)$

Here, Y and O are observed. The parameters to be estimated are $\theta = \{\alpha, \beta, \sigma_\epsilon, \sigma_\omega\}$. Following earlier approaches, we augment the volatility process to the parameter space. The joint posterior distribution of $p(\theta, V|Y, O)$ is proportional to the product of the four conditional distributions:

$$p(V, \theta|Y, O) \propto p(O|Y, V, \theta)p(Y|V, \theta)p(V|\theta)p(\theta).$$

From this joint posterior, the marginal $p(\theta|Y, O)$ can be used to make inferences

about model parameters, and marginal distribution $p(V|Y,O)$ provides the solution to the smoothing problem of inferring about the unobserved volatilities.

As can be seen, both return and option prices are used to make inference on V. The marginal distribution of $p(V|Y,O)$ combines both the historical and implied volatility.

Direct sampling of $p(\theta, V|Y,O)$ is not possible. Instead, we sample the joint posterior $p(\omega, V|Y,O)$ indirectly by iterating through

- $p(\theta|V,Y,O)$
- $p(V|Y,O,\theta)$

The first conditional distribution $p(\omega|v,Y,O)$ can be written as

$$\begin{aligned} p(\theta|V,Y,O) \quad &\propto \quad \prod_{t=1}^{N} p(o_t|y_t,v_t,\sigma_\theta)p(\sigma_\theta) \\ & \qquad \prod_{t=1}^{N} p(y_t|v_t,\alpha)p(\alpha) \prod_{t=1}^{N} p(v_t|\alpha,\sigma_\epsilon)p(\alpha)p(\sigma_\epsilon). \end{aligned}$$

Assuming the standard reference prior on linear models, $p(\theta|V,Y,O)$ is easy to sample from when the three innovations are independent. A technical difficulty is the sampling of $p(o_t|v_t,s_t,\sigma_\omega)$, a Normal distribution with mean equal to $f(S_t,V_t) = E[BS(\bar{V}_t|V_t])]$. There is no analytical solution to this integration since the conditional distribution $p(\bar{v}_t|v_t)$ is unknown. Monte Carlo simulation of this integral within each draw of the Gibbs sampling will be very time consuming. This problem will occur in the sampling of $p(V|Y,O,\theta)$ as well, so we will discuss the solution of it later.

The second conditional distribution is more difficult to sample from. It is not possible to apply the Multi-move Gibbs Sampler and sample the entire vector of V at once because of the non-linear and highly complex structure of $f(S_t,V_t)$. Instead, we decompose the joint distribution into a set of conditionals $p(v_t|v_{-t},o_t,,y_t,\theta)$ where

v_{-t} denotes the rest of the v vector other than v_t.

From model(5), each v_t depends on only the adjacent v_{t-1} and v_{t+1}. Thus, $p(v_t|v_{-t}, o_t, , y_t, \theta)$ can be written as

$$p(v_t|v_{-t}, o_t, , y_t, \theta) \propto p(v_t|v_{t-1}, v_{t+1}, o_t, , y_t, \theta) \quad (3.17)$$

$$\propto p(o_t|v_t, y_t, \sigma_\omega)p(y_t|v_t)p(v_t|v_{t-1}, \alpha, \sigma_\epsilon)p(v_{t+1}|v_t, \alpha, \sigma_\epsilon) \quad (3.18)$$

Equation (3.18) has an unusual form, and it requires the evaluation of the integral $f(\bar{V}|S_t, V_t)$ in $p(o_t|v_t, s_t, \sigma_\omega)$. Since the integral is evaluated within each iteration of the Gibbs sampling n+1 times, Monte Carlo methods of numerical integration to evaluate the integral within each sample of the Gibbs Sampler is computationally prohibitive.

One way to avoid the integration is to approximate the integral is using Taylor expansion. Although an analytical expression for the conditional distribution of $\bar{v}$ is not available, all the moments of $\bar{V}$ can be derived. Therefore, a Taylor series expansion of $f(\bar{V}|S_t, V_t)$ about its expected value $E(\bar{V})$ can be applied. Hull-White (1987) showed that for sufficiently small values of $k = \sigma_\epsilon^2(T-t)$, the series converges very quickly.

Therefore, when $\beta = 0$, (3.17) can be approximated by

$$\begin{aligned} f(S_t, V_t) &= E[BS(\bar{V}|V_t)] \\ &= BS(V_t) + \frac{1/2 S_t\sqrt{T-t}N'(d_1)(d_1d_2-1)}{4\sigma^3}\frac{2\sigma^4(e^k-k-1)}{k^2} - \sigma^4\} \\ &+ \frac{S\sqrt{T-t}N'(d_1)[(d_1d_2-3)(d_1d_2-1)-(d_1^2+d_2^2)]}{8\sigma^5} \\ &* \sigma^6\frac{e^{3k}-(9+18k)e^k+(8+24k+18k^2+6k^3)}{3k^3} + ... \end{aligned}$$

where $k = \sigma_\epsilon^2(T-t)$.

The choice of $\beta = 0$ is justified on the grounds that, for any nonzero β, options of different maturities would exhibit markedly different implied volatilities. Since this is never observed empirically, we must conclude that β is at least close to zero.

This approximation allows us to evaluate the integral at a low cost. Hull and White (1987) compared Monte Carlo simulation results and Taylor Series approximation results for the integral for a 3-month option with $\beta = 0$, $\alpha = 1$ and $\sigma_\epsilon^2 = 0.1$. It can be seen that for short-term at-the-money and out-of-the-money options, this approximation is reasonably well.

In (3.18), the conditional posterior $p(v_t|v_{-t}, o_t, , y_t, \omega)$ is a product of four components: $p(o_t|y_t, v_t)$, $p(y_t|v_t)$, $p(v_{t+1}|v_t)$ and $p(v_t|v_{t-1})$. The last two terms are log-normal distributions and can be approximated and dominated by an inverse Gamma distribution similar to what is used for the Single-Move Gibbs Sampler for the Log-AR(1) model. The inverse Gamma distribution is a good proposal distribution for the Metropolis because it has a fatter tail to the right and therefore can dominate the log-normal distribution. The first two terms in (9) are not of any known distribution form. But $p(y_t|v_t)$ can be evaluated easily and $p(o_t|y_t, v_t)$ can be approximated by the the above series. When the posterior can be written as the product of two distributions, $\pi(t) \propto \psi(t)h(t)$, where $\psi(t)$ can be easily sampled from and $h(t)$ can be easily evaluated, then $\psi(t)$ is a natural envelop for $\pi(t)$ since $h(t)$ is always less than 1. We can use $\psi(t)$ as the proposal distribution to generate samples since $\psi(t)$ will dominate $\pi(t)$. Then the acceptance ratio for the Metropolis distribution is simply

$$\alpha(x, y) = min\{1, \frac{p(y)}{p(x)}\}.$$

See Chib and Greenberg (1993) for a discription of this method.

In summary, the Gibbs Sampler is constructed as follows:

- Sample $p(\omega|V, Y, O)$ by

$$
\begin{aligned}
p(\omega|V,Y,O) &\propto \prod p(o_t|v_t, s_t, \sigma_\omega)p(\sigma_\omega) \\
&\propto \prod p(y_t|v_t,\alpha)p(\alpha)\prod p(v_t|\alpha,\sigma_\epsilon)p(\alpha)p(\sigma_\epsilon)
\end{aligned}
$$

• Sample the joint posterior $p(V|Y,O,\omega)$ by iterating through each individual chain, i.e.,

$$
\begin{aligned}
p(v_t|v_{-t},o_t,,y_t,\omega) &\propto p(v_t|v_{t-1},v_{t+1},o_t,,y_t,\omega), \\
&\propto p(o_t|v_t,y_t,\sigma_\omega)p(y_t|v_t)p(v_t|v_{t-1},\alpha,\sigma_\epsilon)p(v_{t+1}|v_t,\alpha,\sigma_\epsilon).
\end{aligned}
$$

3.3.3 Model Testing

One important use of Stochastic Volatility models is to infer the unobserved conditional volatility processes both within the sample (smoothing) and out of the sample (forecasting). Inference on the volatility can then be used to price and forecast options. In our model, this goal can be achieved in a natural way under the MCMC framework. Past return and option prices contribute to the forecast of volatility and in turn to the forecast of future option prices.

If the option price does follow the stochastic volatility specified, both past behavior and future expectation of price movement will help contribute to the estimation and forecasting of volatility. For in-sample smoothing, this will result in a more concentrated posterior distribution of the volatility than the posterior distribution of volatility from a model for historical or implied volatility only. This can be used to test the hypothesis of the existance of the stochastic volatility. For out-of-sample prediction, this will result in a more accurate forecasting of option prices and hence a higher sharp ratio for the forecasting model.

We have shown how to conduct Bayesian analysis on option pricing theory under MCMC sampling technique when the underlying volatility follows the Hull and White (1987) stochastic diffusion process. This unified approach has obvious theoretical advantage over existing model testing procedures (e.g. option pricing with GARCH type models rely on asymptotic approximations to the continuous time diffusion process; classical analysis on option pricing with stochastic volatility requires an approximate non-linear filter to volatility smoothing.) Bayesian analysis combines option pricing and historical volatility under one model framework and conduct finite sample inference on volatility in a coherent way using MCMC sampling technique.

3.3.4 Extension to the Log-AR(1) Model

In the above section, we illustrated how to combine historical and implied volatilities under one model framework and testing option pricing theory with stochastic volatility using the MCMC simulation technique. We followed the Hull-White (1987) diffusion model for the return and volatility in the analysis. We have illustrated the theoretical advantage of our approach to test option pricing theory with stochastic volatility. The Hull-White model, however, does not incorporate the autoregressive effect commonly observed in financial series. Therefore, forecasting performance of option prices using this model may be limited.

As developed in earlier chapters, we have seen the Log-AR(1) model has the potential to be a successful empirical model for the forecast of historical volatility. It is therefore natural to consider the Log-AR(1) model as the candidate model for historical volatility. Unfortunately, no option pricing formula has been developed when the underlying volatility process follows a Log-AR(1) process.

To understand the difficulty of option pricing with Log-AR(1) stochastic volatility model, we briefly discuss the problem of option pricing with stochastic volatility

Table 3.1: Examples of Continuous-Time Stochastic Volatility Processes

Hull and White	$d\sigma_t^2 = \sigma_t^2(adt + bdz_t)$
Scott	$d\sigma_t = a(b - \sigma_t)dt + cdz_t$
Wiggins	$dlog\sigma_t = a(b - log\sigma_t)dt + cdz_t$
Johnson and Shanno	$d\sigma_t = a\sigma_t dt + b\sigma_t^c dz_t$
Melino and Turnbull	$dlog\sigma_t^2 = a(b - log\sigma_t^2)dt + cdz_t$
Heston	$d\sigma_t^2 = a(b - \sigma_t^2)dt + c\sigma_t dz_t$

in general. Two major difficulties exist in option pricing problem with stochastic volatility. The first is the estimation of unknown parameters in the model. We have developed a reasonably efficient simulation approach to tackle this problem in the previous chapters. The second difficulty in option pricing problem is, like option pricing with stochastic interest rate, option pricing with stochastic volatility does not admit closed-form solutions. A large literature exists on deriving option pricing formula under stochastic volatility. Currently, almost all published continuous-time stochastic volatility models assume a geometric Brownian motion process for the return process. But different models have been suggested for the volatility process. Table 3.1 lists some examples of continuous-time processes for the volatility. As can been, Melino-Turnbull applies a Log-AR(1) model for currency options.

Various approaches have been proposed to price options under the above models. Hull-White derived option price as the average of extended Black-Shole Prices and derived a power series approximation to the expectation. Wiggins (1987) proposed finite differences "hopscotch" numerical methods with logarithmically transformed state variables to solve the partial differential equation. Stein and Stein (1993) and Heston (1993) developed analytic approaches based on Fourier inversion methods. Of these approaches, the numerical solution of the PDE is the most general but, unfortunately, it is also the most computer intensive. The Hull-White and Stein and Stein approaches depend on the distribution of the average stochastic volatility and require instantaneous correlation between increments in state variables to be

zero. Heston applies the Fourier method more generally for nonzero correlation of state variables. Among these methods, the Hull-White power series approximation remains the most tractable one but it is only valid for their diffusion process. Melino-Turnbull used Monte Carlo simulation to price options under the Log-AR(1) model.

It can be seen that analytical approximation to the option pricing formula is only available for limited processes of the volatility. Numerical approximation of option price with our model framework, however, will be computationally prohibitive since the integral is going to be evaluated n times within each iteration of the Gibbs Sampler where n is the number of observations. An efficient option pricing formula is needed to conduct efficient Bayesian analysis on option pricing when the underlying volatility follows a Log-AR(1) process.

Chapter 4

Model Comparison

In Chapter 2, the in-sample model diagnostic results indicate the Log-AR(1) Stochastic Volatility Model fits real exchange rate data well. According to research results from J.P. Morgan, the degree of kurtosis generated by the EGARCH model is smaller than that from the Stochastic Volatility Model which is closer to what is observed in the market.

In this chapter, we develop a model comparison method to compare the Stochastic Volatility model with EGARCH model using option prices. We first study the Smile effect of volatility generated by the market using the Log-AR(1) Model. We then discuss model comparison for financial forecasting models and develop a method to compare the Stochastic Volatility model and EGARCH model. Finally, we discuss option pricing problem with stochastic volatility which is an unsolved problem for the Log-AR(1) Stochastic Volatility Model.

4.1 The Smile Effect

Before we compare the forecasting performance of the Stochastic Volatility Model and EGARCH model, we explore some interesting features of the Stochastic Volatility Model.

First, the posterior distribution of volatility has a log-normal shape with right skewed tails. The skewness statistics for the DEM/USD data is 0.86, the kurtosis statistics is 4.43.

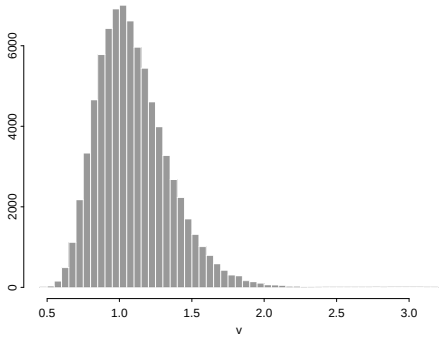

Figure 4.1: *Histogram of the posterior distribution of the last day's volatility*

This is not surprising since the model assumes the volatility follows a log-normal process. This implies if the model is correct, then the traditional inference on volatility using asymptotic normal approximations can be very misleading since it loses the asymmetric information in the distribution of the volatility.

Second, we study the Smile effect generated by the Stochastic Volatility Model. The Smile effect is a very important real-world measure of the deviation of market option prices from Black-Sholes option pricing theory. It is calculated by inverting the volatility parameter from the Black-Sholes option pricing formula after a market price of an option is observed. As stated in the previous chapter, this implied volatility is called a Black-Sholes implied volatility. There are two components of the Black-Sholes implied volatility: one is the volatility smile and the other is the term structure of the volatility.

Volatility smile is the way in which implied volatility varies with strike price for options of a fixed time to expiration. It is usually plotted as the graph of strike price versus implied volatility.

Term structure for volatility describes the way at-the-money implied volatility varies with time until expiration. Thus, to look at the term structure of volatility, we graph time until expiration versus implied volatility for all of the at-the-money

options on the underlying asset.

The volatility smile effect refers to the phenomena that the graph of strike price versus implied volatility has a U-shape. The implied volatility is relatively higher for out-of-the-money options and relatively lower for in-the-money options. This effect is more obvious as time to expiration increases and the degree in which the option is in or out of the money increases. For some options, this smile is also skewed. This effect violates the constant volatility assumption of the Black-Sholes option pricing formula. If we assume Black-Sholes formula price at-the-money option correctly, the Smile effect indicates Black-Sholes formula over-prices out-of-the-money options and under-prices in-the-money options. Since the market crash of October 1987, the Smile effect for most world equity markets has become more pronounced. This has immediate impact on the way options are priced and therefore attracts the attention of researches on this topic. However, as of today, no sound theoretical explanation is accepted for the Smile effect. Of all the competing theories, the stochastic volatility theory is the most popular one. Simulation studies by Hull and White (1986) have shown that stochastic volatility model can generate the volatility Smile effect. Yet as to my knowledge, no one has applied a stochastic volatility model on real market data and study the volatility Smile effect generated from the real data under the model. From our forecasts of volatility using the Log-AR(1) model, we plotted the implied volatility versus different strike prices. The Smile effect is very obvious, which is close to what is observed in the market. This is much bigger than the Smile effect generated from EGARCH model by research results from J.P. Morgan (from personal communication with Dr. Jordan Drachman). The Smile effect generated from our model is consistent with simulation results from Hull and White (1986). This confirms the Smile effect can be effectively explained by the Stochastic Volatility Model for real market data. And this can be considered an alternative way to check the in-sample

adequacy of the Log-AR(1) model.

4.2 Comparing Financial Forecasting Models

We are encouraged by these findings to investigate the forecasting power of the Stochastic Volatility Model and compare with that of the EGARCH model.

Model comparison in general is a very difficult and controversial task. Traditional likelihoods ratio tests of model comparison enable comparison between models only in the nested case where there is an unambiguous null hypothesis. Selection is based on an asymptotic χ^2 approximation (deviance difference) which is inapplicable to models whose number of parameters increases to infinity as the number of observation increases. Most distressingly, classical theory offers little for comparison of non-nested models. See Neyman and Scott (1948) for an overview of classical approaches.

The Bayesian perspective offers a formal decision-based approach for model selection. The Bayesian approach utilizes the marginal distribution of the data $f(Y)$ to assess model performance. Regardless of a model, $f(Y)$ is a density over the space of observables which can be compared with what was actually observed. If Y_{obs} denotes the actual observations and $f(Y|M_i)$ denotes the marginal density under model M_i, $i = 1, 2$ the Bayes Factor

$$BF = \frac{f(Y_{obs}|M_1)}{f(Y_{obs}|M_2)}$$

provides the relative weight of evidence for model M_1 compared to model M_2. The Bayes factor arises formally as the ratio of the posterior odds for M_1 versus M_2 to the prior odds for M_1 versus M_2, see Jeffreys (1961). A table elaborating a rough calibration of the Bayes factor has been proposed by Raftery (1995). A problem with

the Bayes factor is that for many models, at least some part of the prior specifications is vague so that $f(\theta)$ is improper. This makes it difficult to calibrate the Bayes factor.

In the context of financial forecasting models, the $0-1$ loss function adopted by the Bayes factor is not appropriate. Instead, we have a very clear utility function, the Profit and Loss (P/L) generated by the model over a period of time. The model which generates a higher profit over a long period of time will be considered a better forecasting model. In this section, we are going to use the profit and loss generated by the model as the model selection criteria for the comparison of the Stochastic Volatility Model and EGARCH model.

There are many ways in applying a volatility forecasting model in trading and evaluating its P/L. Therefore, the P/L generated by the model will depend on the specific trading strategy applied. Different trading strategies may result in different back-testing results for the P/L. This can be considered as a limitation of our approach. But to build a quantitative trading system and apply it in real trading, we need a consistant and systematic trading strategy to evaluate the performance of a quantitative model. So this approach is considered adequate for our purpose. In the mean time, the need to carefully study the in-sample performance of a forecasting model becomes important since properties from the in-sample model adequacy checking may reveal some important features in the data which may lead to the development of a corresponding trading strategy. For example, if we conclude that the Stochastic Volatiliy model captures volatility smile effect relatively better, we may trade in strips and compare the P/L from trading strips of the two models.

Now we design a trading strategy to compare the Stochastic Volatility Model and EGARCH model. Since volatility is unobserved and not traded, we can't compare the P/L generated by forecasting volatility. But since volatility is the only unknown parameter in the pricing of options, we can use volatility forecasts from both models

Table 4.1: Payoff of a straddle

Range of Stock Price	Payoff from call	Payoff from put	Total payoff
$S_T \leq X$	0	$X - S_T$	$X - S_T$
$S_T > X$	$S_T - X$	0	$S_T - X$

to price options and compare the P/L from trading options.

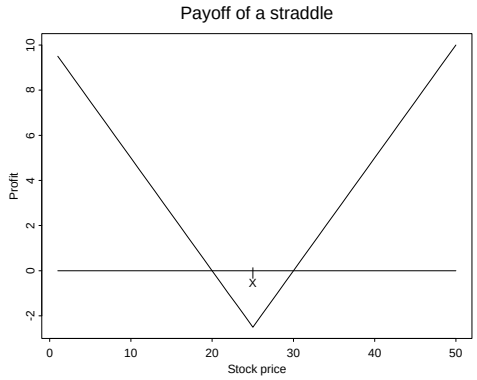

Figure 4.2: *Payoff of a straddle*

There are many option trading strategies. For example, writing naked positions, covered positions, spreads, straddles, strips, etc. For proprietary trading purpose, if we believe that we have a good forecasting model on volatility, we should select a trading strategy that bets on volatility. Trading straddles is such a strategy. A straddle is a combination of buying a call and put option with the same strike price and expiration data. The profit pattern is shown in Figure 4.2. The strike price is denoted by X. If the stock price is close to this strike price at expiration time of the option, the straddle leads to a loss. However, if there is a large move in either direction of the stock price, a significant profit will result. The payoff from a straddle is calculated in Table 4.1.

The straddle in Figure 4.2 is also called a bottom straddle or straddle purchase. A top straddle or straddle write is the reverse position. It is created by selling a call and a put with the same exercise price and expiration data. It is a highly risky strategy. If the stock price on the expiration date is close to the strike price, it leads

to a significant profit. However, the loss arising from a large move in either direction is unlimited. One example of the misuse of the selling of straddles is the collapse of Barings Bank, see the report by Martin (1996).

So in practice we need to delta hedge the position when buying or selling a straddle. A delta of a derivative, $\triangle$, is defined as the rate of change of its price with respect to the price of the underlying asset. See, Figure 4.3. It is the slope of the curve that relates to the derivative's price to the underlying asset price. An approximation to the slope is defined by

$$\triangle = \frac{\triangle C}{\triangle S}$$

where $\triangle S$ is a small change in the price of the underlying asset and $\triangle C$ is the corresponding change in the option price. It is the number of units of the stock needed to be held for each option shorted to create a riskless hedge. The construction of such a riskless hedge is sometimes referred to as delta hedging. When the delta of the position in the underlying asset offsets the delta of the option position, the delta is equal to zero and the position is called delta neutral. By delta hedging the position, the downside risk of selling an option is eliminated.

After a trading strategy is selected, a criteria for entering and leaving a position needs to be constructed. The basic principle of trading is to buy if the market price is undervalued and sell if the market price is overvalued. This means we should enter a position (buy an option) if the market price is below the forecasted price and leave a position (sell an option) if the market price is above the forecasted price. A quantitative approach to determine the entering or leaving of a position is to use a confidence interval from the posterior distribution of the forecasted option prices. If the market observed value falls out of the confidence interval of the posterior distribution, we enter or leave a position. Because of the time value of money, we

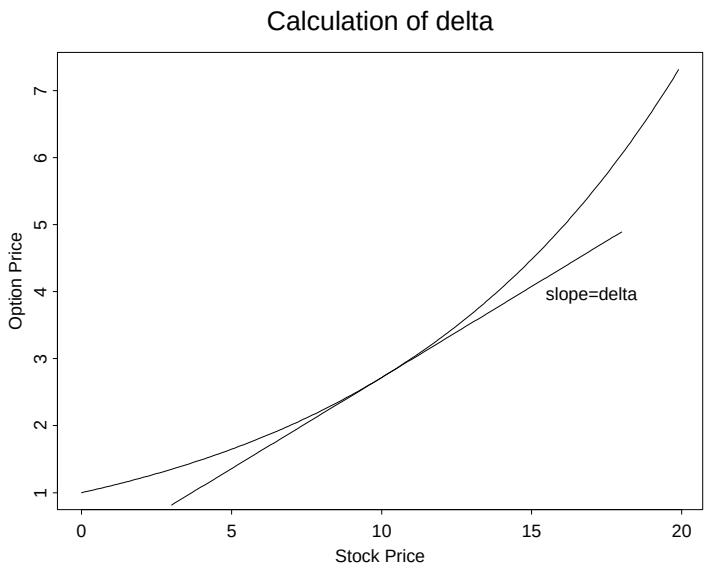

Figure 4.3: *Calculation of delta*

want to be more confident when entering a hedged position than leaving a hedged position. Therefore, a 95% confidence interval for the option price is used for entering a hedged position (buy an option) and a 50% confidence interval is used for leaving a hedged position (sell an option).

Another issue is the volumn of each trade. One way is to assume equal volumn of each trades. A better approach would be to relate the volumn of the trade with the percentile of the observed option price on the posterior distribution of the forecasting option prices. This will involve the concept of Expected Sharp Ratio. The expected Sharp Ratio is defined as the ratio of the expected return devided by the variance of the expected return. For each unit of trade, the Expected Sharp Ratio is defined as

$$R = \frac{E(r)}{\sqrt{Var(r)}}$$

Similarly, for n units of trades, the Expected Sharp Ratio is defined as $R = \frac{nE(r)}{n^2\sqrt{Var(r)}} = \frac{E(r)}{n\sqrt{Var(r)}}$. So, the Expected Sharp Ratio for n trades is a function of the

trading volumn n. This ratio measures the risk adjusted return and is a commonly used measure for the performance of a forecasting model. A ratio bigger than zero for a forecasting model indicates the model is making a profit. Similarly, an expected ratio bigger than zero for a trade means an expected postive cash flow. Unlike the return-based criteria which uses the confidence interval of the forecasting distribution of option prices to determine for entering or leaving a position, the Expected Sharp Ratio distinguishes the volumn of each trade.

In summary, the specific trading strategy we select is to trade delta hedged straddles using the forecasting distribution of option prices. The criteria for enterting or leaving a position is the Expected Sharp Ratio being bigger than zero. We compare the P/L performance of the two models in trading delta hedged straddles over a certain historical period. The model with higher profit will be considered a better forecasting model.

4.3 Bayesian Option Pricing with Stochastic Volatility

Another issue in the back-testing procedure is an option pricing formula with stochastic volatility. This is the major obstacle in the implementation of the comparison of out-of-sample forecasting performance of the Log-AR(1) Stochastic Volatility Model and EGARCH model. As stated in the previous chapter, no analytical approximation of options prices under the Log-AR(1) Stochastic Volatility Model has been developed.

Option pricing with stochastic volatility in general is a very difficult task and is beyond the scope of this dissertation. We will discuss the major developments in this area and discuss numerical approximation approaches to the option prices. We also discuss Bayesian analysis of option prices.

A vast literature exists on the development of an option pricing formula with stochastic volatility. Hull and White (1986) derived an option pricing formula when the underlying volatility follows a diffusion process with constant drift. Heston (1990) developed a closed-form solution to allow a more general stochastic process for the volatility. The volatility process in Heston's model is

$$d\sigma_t^2 = a(b - \sigma_t^2)dt + c\sigma_t dz_t,$$

Yet no analytical work on the pricing formula for options when the underlying volatility follows the Log-autoregressive process has been done. The closest formula avaliable for the Log-AR(1) process is the general conclusion of Hull and White (1986) on representing option price as the average of the extended Black-Sholes price which is true when the underlying volatility follows a Log-autoregressive process, e.g.,

$$O_t = \int BS(\bar{V}_t)h(\bar{V}_t|V_t)d\bar{V}_t = E[BS(\bar{V}_t|V_t)],$$

where $\bar{V}_t = \frac{1}{T-t}\int_t^T V_i di$.

Therefore, one approach to price options using the Log-AR(1) model is to do simulation over the Hull and White option price integral. Comparing with simulation over the original definition of option prices, simulate using the average of the extended Black-Sholes price formula only requires the simulation of the volatility process. This cuts the simulation time in half comparing to simulation of both the volatility process and return process using the original definition of option prices.

Another issue we discuss is the estimation of option prices. Most (and to my best knowledge, all) current literature uses point estimates for option prices. A confidence interval is constructed assuming asymptotic normal distributions for the option price for error estimate. In the context of Bayesian analysis, a posterior

distribution can be constructed using the distribution of v_t. From either the Black-Sholes model with constant volatility assumption or Hull-White model with stochastic volatility, the option price is a funtion of the current volatility, $O = f(v_t)$. From the Gibbs Sampler, a posterior distribution of v_t is avaliable. Therefore, simulation-based option prices can be obtained using each sample from the posterior distribution. And an empirical posterior distribution of the option price can be obtained from the posterior distribuion of v_t. This posterior distribution will contain more information than a point estimate of the option price using asymptotic normal assumption. It is well known that returns of stock price, exchange rates have fat tail and skewed distributions from extensive empirical studies. Yet the distributional property of the option prices are rarely studied. The key issue is computational difficulty which involves double-simulation of the option pricing integral and the posterior distribution of the option prices. Yet studies on the distributional property of option prices will have important applications in option trading and risk management of derivative products.

Chapter 5

Conclusions and Future directions

The goal of this part of the dissertation is to develop simulation methods to estimate the highly difficult Stochastic Volatility Model and evaluate the performance of Stochastic Volatility Models as a forecasting model and compare with the existing popular EGARCH model. Motivated by the work of Jacquier, Polson and Rossi (1994), we develop two efficient MCMC sampling procedures to fit a Log-AR(1) Stochastic Volatility Model. We then develop model diagnostic procedures to check in-sample model adequacy. We apply the model to real exchange rate data series. We found that both MCMC simulation procedure developed in the dissertation works well on real financial data where the correlation of volatility is high. In particular, the convergence of the multi-move Gibbs Sampler is much faster than the Single-move Gibbs Sampler. Our model diagnostic results indicate that the Log-AR(1) Stochastic Volatility Model captures the fat tail property of the data better than the EGARCH model. Also, the Smile effect generated by the Log-AR(1) Stochastic Volatility Model is bigger than that from EGARCH model and is closer to the Smile effect observed in the market. This suggests the forecasting behavior of the stochastic volatility model might be different from that of the EGARCH model.

There are many immediate extensions to the current study. A first immediate follow up would be the comparison of the out-of-sample forecasting performance of the Stochastic Volatility Model and the EGARCH model. As stated in Chapter 4, the most commonly used criteria for selecting financial forecasting models is out-of-sample forecasting performance from backtesting results.

There are many unsolved issues in the back-testing procedure. The first issue

is the derivation of an option pricing formula. Since the loss function used in the comparison of the two volatility forecasting models is the profit and loss generated by the trading of options using the model, an option pricing formula under the Log-AR(1) Stochastic Volatility model needs to be developed. Option pricing under stochastic volatility will involve simulation. Since simulation of option prices will be performed daily for a long period of time, approximations or innovative simulation techniques need to be developed to reduce simulation time and make large scale back-testing feasible.

The second issue involved in the back-testing process is the development of a Bayesian trading strategy. The common practice in the estimation of option prices in both academic research and in practice is to conduct point estimates of option prices. However, it is clear from both the Black-Shole's model and the Hull and White option pricing formula with stochastic volatility that option price is a function of the volatility. Since the volatility has a distribution which is usually skewed, use of a point estimation of option price as a summary of such distribution will definitely lose information. A trading strategy using the distribution of the option price will reduce the risk of selling not heavily over-priced options and buying not heavily under-priced options. Since such distribution of option prices will be simulation-based, an efficient option pricing formula and innovative simulation techniques need to be developed so that practical use of the Bayesian idea is feasible.

The third immediate extension to the current study is the extension to the correlated stochastic volatility model. Such a model will account for the skewness effect commonly observed in the stock market and the exchange rates from the emerging markets. As discussed in Chapter 3, the introduction of the correlation destroys the variance-covariance structure. As a result, the Multi-move Gibbs Sampler can no longer be applied. Therefore, we again face the problem of slow convergence of

the Single-move Gibbs Sampler. Innovations in either model modification or MCMC sampler need to be developed to speed up convergence.

The fourth extension to the current study is the development of Multivariate Stochastic Volatility Model. This is of use in portfolio management and cross-sectional analysis. Multivariate analysis of volatility in general is difficult and rather new. Computational difficulties are the major obstacles. Developments in multivariate GARCH type models have experienced difficult identification problems. Earlier work includes Kraft and Engle (1982), Bollerslev, Engle and Wooldridge (1988), Bollerslev (1990). These models are either extremely unparsimonious or quite tightly constrained. Another important extension in the GARCH literature is the factor type GARCH models, following the work of Engle, Ng and Rothschild (1990) and Diebold and Nerlove (1989) which is refined by King, Sentana and Wadhwani (1994). Such models have the potential to improve the parsimony problem.

The corresponding multivariate extensions to the Stochastic Volatility Models are easier than the multivariate GARCH models. A representative work is by Harvey, Ruiz and Shephard (1994) who used Quasi-likelihood Kalman filtering technique to fit a model of the form

$$y_{it} = \epsilon_{it} e^{h_{it}/2},$$

where $h_t = (h_{1t}, ...h_{NT})'$ follows a multivariate random walk and $\epsilon_t = (\epsilon_{1t}, ...\epsilon_{NT})' \sim NID(0, \Sigma_\epsilon)$. Other work on this model includes Mahieu and Schotman (1994) and Jacquier, Polson and Rossi (1995) who applied a MCMC sampler to the model. Another promising direction in the multivariate extensions to the Stochastic Volatility Models is the factor type Stochastic Volatility Models which can capture the long memory property in the volatility process.

The fifth extension to the current study is to implement the model developed in

Chapter 3 which combines implied volatility and historical volatility under one model framework. As discussed in Chapter 3, if the Log-AR(1) Stochastic Volatility Model does capture the behavior of the historical volatility, then a model which utilize both historical information from past return data and future expectation incorporated in option prices based on correct characterization of the historical movement of return will have great potential to provide better forecasting of future volatility and option prices than any single model which models only the historical volatility or the implied volatility. On the other hand, such a model can provide useful empirical evidence for the hypothesis of the existence of stochastic volatility. For example, if the volatility does follow the Log-AR(1) structure specified in the model, the addition of option prices information will result in a more concentrated posterior distribution for the volatility than the posterior distribution obtained from a model for historical or implied volatility only. The main issue in the implementation of such a model is again an efficient option pricing formula and efficient simulation technique for simulating the option prices. This is especially crucial since the simulation of option prices needs to be done n times within each iteration of the Gibbs Sampler.

Other extensions in the Stochastic Volatility Models include the heavy tail Stochastic Volatility Models, the Volatility-in-mean Stochastic Volatility Model which incorporates volatility in the mean part of the return process, etc.

Research on stochastic volatility is growing rapidly. With the ever increasing computer power and development of innovative estimation methods, we are ready to explore many interesting topics in theoretical finance. For example, how to make inference on continuous-time models in theoretical finance using discrete-time data especially high frequency intra-day data; the relationship between stochastic volatility and trading volume (Gallant, Hsieh and Tauchen,1991, Anderson,1995); understanding the cause of stochastic volatility using time deformation (Ghysels, Gourieroux

and Jasiak,1996), etc. On the other hand, Stochastic Volatility Models provide a good testing ground for the development of new nonlinear and non-Gaussian time series techniques for model estimation, model diagnostic and model comparison.

In addition, Bayesian MCMC methods provides excellent framework for the estimation of other sophisticated hierarchical models in finance where a latent process has sound economic intuition. For example, the jump-diffusion models used in option pricing theory and the Regime-Switching models used in interest rate forecasting, etc. These are going to be my continuing research topics at Prudential Securities.

Chapter 6

Introduction

The problem of spectral inference in time sampled data is very old. The first significant advance in frequency estimation occured in the early 20th century when two separate methods of analyzing the problem came into being: the use of probability theory and the use of Fourier transform. Two prominent limitations of the Fourier transform approach exist: Fourier transform can not distinguish multiple signals effectively, and implicit windowing of the data occurs when the Fast Fourier transform is performed. The ''leakage"from the main lobe of a spectral response causes the distorting and obscuring of other spectral responses. On the other hand, use of probability theory under a statistical model permits one to incorporate one's knowledge about the process from which the data samples are taken (e.g. the number of frequencies in the signal). By adding reasonable assumptions to the model, one can describe the nature of the process outside the measurement interval more precisely. Another advantage of using statistical models is the elimination of the windowing effect when processed with the Fast Fourier Transform.

In over 100 years since the introductory of spectral analysis and probability theory, no particular connection between the two has been noted. That the two methods could be very related was first introduced when Jaynes (1983) derived the periodogram directly from probability theory and demonstrated it to be a sufficient statistic for the inference of a single stationary frequency in a time series with a normally distributed noise. Direct Bayesian inference of the frequency was first developed by Bretthorst (1987) under an harmonic model framework. Bretthorst provided comparison between Bayesian posterior analysis of the frequency and the

traditional Fourier transformation. However, the approximations used by Bretthorst for estimation of frequency are not accurate when multiple close frequencies exist, the case when Fouriour transform fails to work. Furthermore, these approximations can not be generalized to more complex models for the frequency. Therefore, although theoretical framework for Bayesian analysis of frequency is avaliable, development of methods for accurate estimation of the frequency is still an essentially untouched area of research.

In this part of the dissertation, we develop a unified approach for frequency estimation using the Gibbs Sampler under the single-frequency and multi-frequency harmonic model framework. The focus of the analysis is on the computation of the posterior distribution of the frequency. I develop efficient procedures to tackle the challenging problem of sampling highly multi-modal posterior distribution for the frequency. The method is then applied to study some EEG and Oxygen isotope data. Motivated by the study in Oxygen data and other geophysical time series, I study the impact of uncertain timing in the estimation of frequency and develop methods for the estimation of frequency and time given observed time series data.

This part of the dissertation is organized as follows. Chapter 7 develops a Metropolis algorithm to sample the posterior distribution of frequency under the single-frequency harmonic model. Chapter 8 extends the methods to a multi-frequency harmonic model and improves the methods for the high dimensional cases. Chapter 9 discusses the impact of uncertain timing in frequency estimation and develops a unified Gibbs Sampler approach for frequency estimation under uncertain timing. This general framework can be generalized to more sophisticated models for frequency domain time series analysis. Chapter 10 discusses future directions on the study of spectral analysis. Illustration of the methods by analyzing the EEG data and Oxygen data are presented at the end of each chapter.

Chapter 7

A Single-frequency Harmonic Model

Bretthorst (1987) first adopted Bayesian posterior distribution analysis for frequency estimation. However, Bretthorst was only able to conduct approximate estimation to the Bayesian posterior distribution of the frequency. Such approximations work well only when the assumptions of Fourier transformation hold. Therefore, as discussed in the previous chapter, although Bayesian posterior analysis of the frequency is established under more general conditions than Fourier transformation, accurate posterior estimation procedures are needed.

In this chapter, we follow the basic harmonic model developed by Bretthorst (1987). We develop a general approach for frequency estimation using the Metropolis algorithm to sample from posterior distribution of the frequency and a Gibbs sampler to conduct inference on other model parameters. The Gibbs sampler is used to conduct exact finite-sample inference for the model parameters especially for the frequency and can be generalized to more complex dynamics for the frequency. I first discuss the harmonic model and the connection between Bayesian posterior distribution for the frequency and traditional Fouriour transformation. Then I propose an efficient simulation procedure to conduct exact finite sample estimation on the frequency. Finally, I apply the model to analyze some EEG data series.

7.1 The Model

Suppose we have a time series y(t) over time t observed at times $\{t_1, ...t_n\}$ to give data $y(t) = y(t_i), \{i = 1...n\}$.

The simplest harmonic regression model is

$$y_t \quad = rcos(\omega t + \phi) + \epsilon_t \quad \epsilon_t \sim N(0, v) \tag{7.1}$$

Here ω is the frequency of the signal, $r > 0$ represents the amplitude $\phi \in (0, 2\pi)$ is the phase, and ϵ_t is a zero mean error process. In this case, $\epsilon_t \sim N(0, v)$ independently.

Rewrite (2.1) as

$$y_t \quad = acos(\omega t) + bsin(\omega t) + \epsilon_t \quad \epsilon_t \sim N(0, v), \tag{7.2}$$

where $a = rcos(\phi), \quad b = rsin(\phi)$ and $r = \sqrt{a^2 + b^2}$.

In the case of equally sampled data, e.g. $t_1 = 1, t_2 = 2, ...,$ w is between $(0, 2\pi)$. For $0 < \omega < \pi$, we get the same model with the sign of b changing to $-b$, at frequency $2\pi - \omega$, so the range of ω can be further restricted to $0 < \omega < \pi$. This implies that $w/2\pi < 0.5$, or $\lambda > 2$. So at least two equally spaced observations are needed to be sampled per cycle.

With arbitrary equally spaced data, e.g. $t_i = i\delta$ for some $\delta > 0$, the range for ω is $0 < \omega < \pi/\delta$, correspondingly, and the wavelength λ has to be bigger than 2δ.

When the time scale is changed, e.g. $u = (t - x)/s$, (7.1) changes to $y_t = rcos((sw)u + (\phi + xw)) + e_t$, with the same amplitude.

Model (7.1) assumes the underlying process as a single cosine wave with fixed frequency, amplitude and phase, plus some white noise with constant variance.

7.2 Posterior Distribution for the Frequency

We are mostly interested in the estimation of the frequency in a harmonic analysis. Maximum likelihood estimates and Least Square estimates of the frequency have been used for many years. See Kay and Marple (1981) for an extensive review. In the Bayesian paradigm, the marginal posterior distribution of the frequency can be obtained by integrating out other parameters, e.g, the amplitude, the phase and the variance of the noise. Details can be found in Bretthorst (1988), as follows.

Assuming fixed ω, rewrite (7.2) as

$$Y = X\beta + \epsilon_t \ \epsilon_t \sim N(0, v)$$

where $X_{n*2} = (cos(\omega t), sin(\omega t)) \ \ \beta = (a, b)'$

Given the reference prior $p(\beta, \sigma^2) \propto 1/\sigma^2$

$$p(Y|X) = p(Y|w) = \int\int p(Y|X, \beta, \sigma^2) p(\beta, \sigma^2) d\beta d\sigma^2.$$

So,

$$p(Y|\omega) \ \ \sim \ \ |X'X|^{-1/2}(1 - (\hat{\beta}'X'X\hat{\beta})/(Y'Y))^{(2-n)/2} \tag{7.3}$$

where $\hat{\beta} = (X'X)^{-1}X'Y$ is the Least Square estimator of β.

The calculation of (7.3) is easy to implement. Nevertheless, when n is reasonably large, ω is not too small, and the data are equally spaced, a useful approximation arises. Using the facts that

$\sum_{i=1}^{n} cos^2(wt_i) \simeq n/2,$

$\sum_{i=1}^{n} sin^2(wt_i) \simeq n/2,$

$\sum_{i=1}^{n} (cos(wt_i) sin(wt_i)) \simeq 0,$

and $(x'x)^{-1} \simeq 2/nI$,

Let

$d^2 = (Y'Y)/n$,

$R(w) = \sum_{i=1}^{n} y_i cos(wt_i)$,

$I(w) = \sum_{i=1}^{n} y_i sin(wt_i)$,

and $C(w) = (R(w)^2 + I(w)^2)/n$.

Then $p(Y|\omega)$ simplifies to

$$\text{p(Y—}\omega) \quad \propto \quad (1 - 2C(\omega)/(nd^2))^{(2-n)/2}.$$

Under a uniform prior $p(\omega) \sim c$, the posterior for the frequency is

$$p(\omega|Y) \propto (1 - 2C(\omega)/(nd^2))^{(2-n)/2} \tag{7.4}$$

Here $C(\omega)$ is called the periodogram, introduced by Schuster 200 years ago. Traditionally, the view is that large $C(w)$ implies important frequencies because $C(w) \propto \hat{a}^2 + \hat{b}^2$ which estimates the theoretical squared amplitude $r^2 = a^2 + b^2$. The posterior for ω in this one-cycle model is a monotone function of $C(w)$ under the above approximation. Now under the pioneer work of Jaynes, the two separate methods of estimating frequencies, probability theory and Fourier transform, were beautifully linked together under the Bayesian paradigm. Furthermore, the posterior distribution greatly improves the resolution of the frequency and eliminates the windowing effect of the Fourier transform.

Remember that, the above approximation is true only when the following assumptions are held, e.g.,

- the number of data n is large;

- there is no very small frequency;
- there is only one frequency in the data;
- the amplitude and phase are constant;
- the data are equally sampled.

When these assumptions are not held, discrete Fourier transformation and the periodogram are no longer well defined, e.g, Fourier transform always interprete the data as a single frequency harmonic model. Yet, the posterior distribution of ω is still perfectly well-defined and gives the correct scale of the estimation of the frequency without windowing the data. Furthermore, assessment of the estimates is avaliable from the posterior distribution of the frequency which neither least squares, maximum likelihood, nor the Fourier transform can give directly.

7.3 Metropolis Sampling of the Posterior Distribution

Inference from the posterior distribution is a challenging problem. The posterior distribution $p(\omega|Y)$ does not have a standard form. As a result, analytical approximation and simulation methods need to be developed to achieve inference in ω.

Bretthorst (1988) adopted an analytical approximation approach to the problem in which he applied a t-distribution approximation to the approximated posterior distribution (7.4). Such appoximation is obviously crude with today's sampling techniques. In addition, it is rather limited since the approximation equation (7.6) holds only when the assumptions of Fourier transforms are held. As discussed in the previous section, when assumptions of Fourier transforms are violated, the approximation of (7.5) using (7.6) is already not true, not to mention the t-distribution approximation to it. So although Bayesian posterior probability approach has shown theoretical advantage under more general assumptions, direct inference from the posterior dis-

tribution was still not avaliable under general conditions.

Sampling the posterior distribution of the frequency is a very challenging problem. A plot of $p(\omega|Y)$ reveals that $p(\omega|Y)$ is typically highly multi-modal. The probability mass is highly concentrated at one or several major peaks which indicates the "important frequencies" in the data. When the signal is strong, the intervals where these main frequencies lie are extremely small comparing to the standard $(0, \pi)$ interval for the frequency. Meanwhile, numerous local modes exisit in other areas of the distribution because of the property of trigometry. This makes it hard to apply standard accept/reject sampling methods since the search of an "envelope" for the posterior distribution is very difficult without knowledge of the location of the important frequencies. A random walk Metroplis sampling with a small variance will tend to be misled by local modes and one with a large variance will result in large rejection rates. We consider an independence chain Metropolis algorithm with a good proposal distribution as the most efficient approach to sample the posterior distribution.

The advantage of an efficient algorithm for the sampling of the frequency parameter is not obvious in the single-frequency harmonic model since the sampling of the frequency is independent of other parameters. But in the following chapters the need for an efficient sampler for the frequency is obvious when we introduce multi-frequency model, time-varying amplitudes and timing uncertainties into the basic harmonic model. Therefore, we spend some effort on developing an efficient Metropolis algorithm for the sampling of frequency.

In an independence chain Metropolis algorithm, a good proposal distribution is crucial for efficiencient sampling. It is expected to mimic the targeting distribution reasonably well, and the tails of the starting distribution should dominate those of the targeting distribution to insure convergence of the Markov chain. Also, it should be fast to compute. Three choices of starting distributions are presented here.

7.3.1 Mode-Based Proposal Distribution

One choice of proposal distribution is a mixture of t-distributions centered at the modes of the targeting distribution. It was proposed by Gelman and Rubin(1992). Such mixture distribution should mimic the shape of the targeting distribution reasonably well, and a t-distribution with small degree of freedom will gurantee dominance at the tail. Although mode-based distributions don't work for every problem, it is a good choice for this univariate exponential like distribution.

As stated earlier, $p(\omega|Y)$ is highly multi-modal and highly concerntrated around the major peaks. Therefore, mode-finding for $p(\omega|Y)$ isn't an easy task. Since the first derivative of $p(\omega|Y)$ is analytically avaliable and with similar structure of $p(\omega|Y)$, Newton's method is known to be the fastest mode searching method. However, Newton's method requires fairly accurate starting values to guarantee correct convergence. Because of the numerous local modes presented in $p(\omega|Y)$, it's very likely for Newton's method to converge to local modes. On the other side, since the posterior is highly concerntrated around the peak, large number of starting points are needed to insure a starting point close to the peak was selected. For equally sampled data, the problem is easier since the frequency is known to be between $(0, \pi)$, and in most practical problems, prior knowledge allows us to further constrain the interval. This allows us to sample thousands of points in the interval $(0, \pi)$ to guarantee a close starting point to be used. The procedure to construct the mixture starting distribution is as follows:

1. Randomly sample a large number of starting points from $(0, \pi)$.

2. Perform Newton's method using each starting point.

3. Evaluate $p(\omega|Y)$ at each local mode. Select K most important ones.

4. Compute the second derivative at each local mode as an estimate of the variance at the mode.

5. Construct a mixture of t-distributions centered at the modes with a small degree of freedom as a starting distribution.

Here, much attempt has been made on obtaining an accurate starting distribution. Since the searching for starting value of Newton's method is needed to be done only once in the single-frequency model, it's worth spending the computing time here to guarantee fast convergence of the Metropolis algorithm.

7.3.2 Kernel Density Proposal Distribution

An alternative way to find a good proposal distribution is to use kernel density estimate as an approximation of the posterior distribution, see West (1992). The idea is as follows, in exploring a univariate posterior density, $p(x)$, in general,

• Draw a sample of size n (usually several thousands points) $x_1....x_n$ from a uniform distribution which covers the interval in which the frequency lies.

• Evaluate the weights $w_j, \{j = 1..n\}$ determined by $\omega_j = p(x_j)/\sum_{j=1}^{n} p(x_j)$

• construct a mixture function

$$p(x) = \sum_{j=1}^{n} \omega_j d(x_j, vh^2)$$

The function $d(x_j, vh^2)$ can be either a Normal distribution or a T-distribution.

The quantity v is the Monte Carlo estimate of posterior variance

$$v = \sum_{j=1}^{n} w_j (x_j - \bar{x})^2$$

where $\bar{x} = \sum_{j=1}^{n} w_j x_j$.

The scaling parameter h is usually chosen as a decreasing function of the sample size n, so that the kernel components are more concerntrated at the locations x_j for larger sample sizes. A traditional specification with normal kernels is

$$h = c/n^{1/(1+4p)}$$

with $c = \{4/(1+2p)\}^{1/(1+4p)}$ although much smaller c can be found for specific models.

This approach is similar to the first one in computing time and accuracy, but is more general and is suitable for problems where an interval within which the parameter lies is unknown.

Once the starting distribution $q(y)$ is determined, Metropolis algorithm proceeds as follows:

Repeat for $j = 1, 2, ...N$

- Generate y from $q(y)$, and generate u from $U(0, 1)$,
- Let $x^{(j+1)} = y$ if $U < \alpha(x^j, y)$; otherwise let $x^{(j+1)} = x^j$.

Return the values $\{x^{n_0+1}, x^{n_0+2}, ...x^N\}$. Here n_0 is the "burn-in" period needed to ensure the chain has passed the transient stage and converged to the targeting density. Since the starting distribution found by the above three ways are fairly accurate, the convergence of the Metropolis algorithm should be fast.

7.4 Estimation of Other Model Parameters

Besides the frequency, other parameters to be estimated in model (7.2) are the amplitudes, and the variance of the noise. The amplitudes are useful in determining the importance of the frequency and the power carried by the signal. The variance of the

noise is a good indication of the signal-to-noise level and can help determine whether one should try a new model for the data.

Bretthorst (1988) estimated the amplitude and the variance assuming the frequency is known since direct analytical forms of the marginal distribution of the amplitude and the variance are unobtainable. To get direct inference from the marginal distribution for the amplitude and the variance, a Gibbs sampler is used.

To generate samples from the marginal distributions for the amplitude and the variance,

Rewrite (2.2) as a linear model

$$Y = X\beta + \epsilon$$

where the design matrix $X_{n*2} = (cos(wt), sin(wt))$, $\beta = (a, b)'$, and ϵ is a $n * n$ diagnoal matrix with v as the variance of each ϵ_t.

The joint posterior is

$$p(a, b, v, \omega|Y) \propto p(\omega|Y)p(v|y, \omega)p(a, b|y, \omega, v)$$

Under the assumption that (a, b, v) are it a priori independent of ω, using standard Bayesian linear model theory, we sample $p(a, b, v|Y, \omega)$ by composition. Using a uniform prior for ω, and a reference prior $p(\beta, \sigma) \propto 1/\sigma^2$, $p(v|y, \omega)$ is an inverse Gamma distribution with shape parameter $(n-2)/2$ and scaled parameter $ns^2/2$. The conditional distribution $p(a, b|y, \omega, v)$ is a bivariate normal distribution.

After breaking the joint posterior into several conditional distributions, the full posterior can be sampled by sequentially simulating as follows:

- Sample ω from

$$p(w|Y) \propto |X'X|^{-1/2}(1 - (\hat{\beta}'X'X\hat{\beta})/(Y'Y))^{(2-n)/2},$$

- Sample v from $p(1/v|Y, w) \sim Ga((n-2)/2, ns^2/2)$

- Sample (a, b) from

$$p(a, b|Y, w, v) \sim N(\hat{\beta}, (X'X)^{-1}v), \quad \hat{\beta} = (X'X)^{-1}X'Y$$

In this sampling scheme, sampling of ω is independent from the sampling of other model parameters. Therefore, the mode-finding procedure in the Metropolis algorithm needs to be done only once.

7.5 Application

One motivating problem of harmonic analysis arises from the collaboration with Duke psychiatrists studying issues of clinical design and efficacy of brain seizure treatments. Electroconvulsive therapy (ECT) is a major tool in brain seizure treatment. EEG monitoring is the primary method of observation on brain activity during ECT. Very long EEG time series, of the order of several tens of thousands of recordings, are available from individuals under varying treatments. Such series usually exhibits quasi-cyclicity among other features. Figure 7.1 shows a plot of a single series of 300 EEG data over time. There are obvious periodicities in the data and the amplitudes vary as well. We assume the data are equally spaced and apply the single frequency harmonic model to the data.

Figure 7.1 also shows a smoothed periodogram estimate of the spectral density function using the default spec.pgram in S-plus, the Schuster's periodogram $C(\omega)$, the Bayesian posterior distribution calculated from the single frequency model as well as the log-posterior distribution. These plots are consistent with each other and indicates a single high frequency at $\omega = .2975$ (wavelength $\lambda = 21.13$) and some minor frequencies.

We estimate the parameters in the model using the sampling procedure described in section 4 of the chapter. First, 5000 samples of the frequency given the data are drawn. Then, samples of the variance given the frequency and the data are drawn

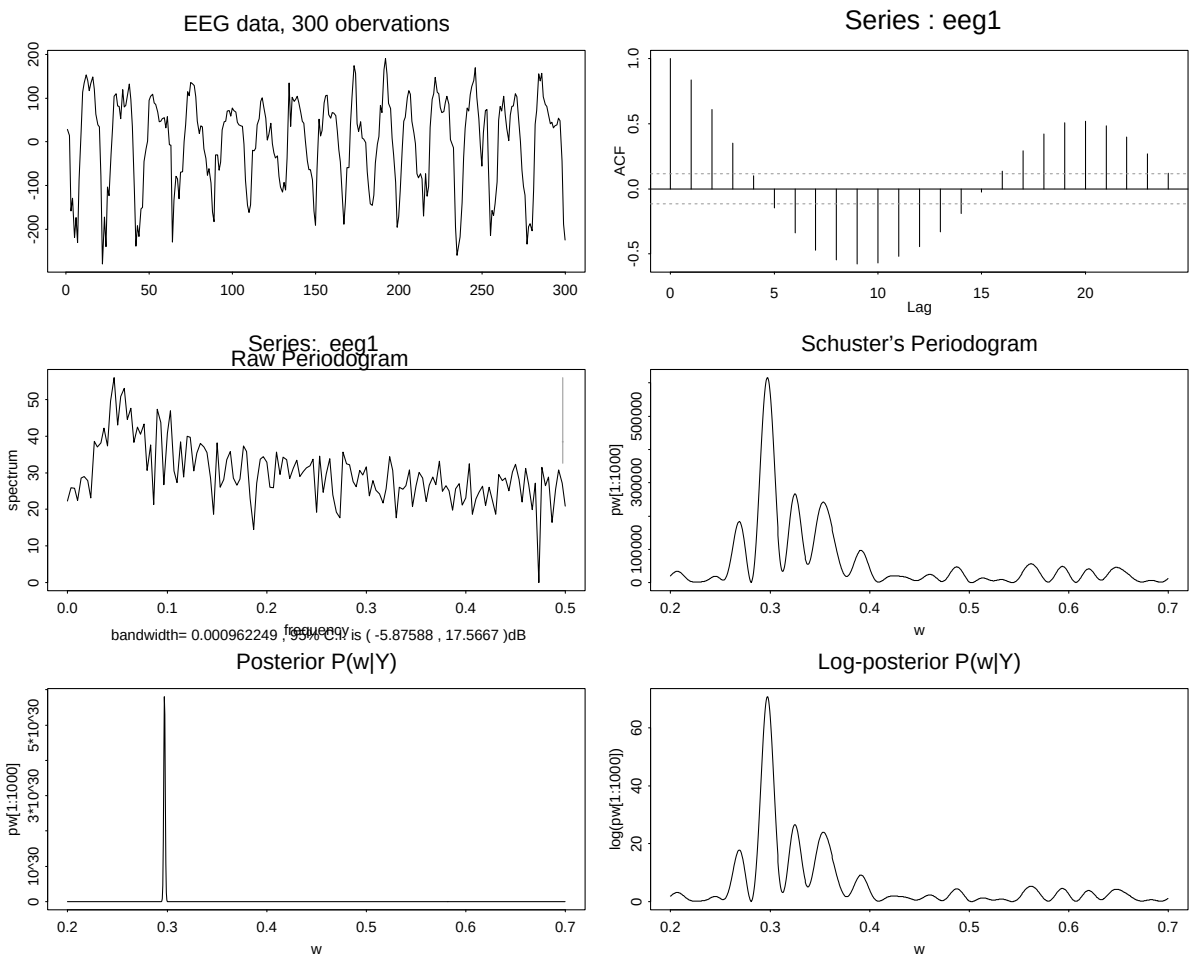

Figure 7.1: *Time series plot of the EEG data. The acf, periodogram , Schuster's periodogram, posterior distribution and log-posterior distribution of the frequency.*

and finally samples of the amplitudes given the frequency and the variance are drawn.

An independence chain Metropolis algorithm is used to sample from the posterior distribution for the frequency. The mode-finding procedure defined in 7.1 was used to construct the mixture of t-distribution as the candidate generating distribution. For this data set, there is only one significant mode in the posterior distribution, so the candidate generating distribution is a t-distribution centered at the mode. The degree of freedom of the t-distribution was chosen to be 3 to insure it's over-dispersed. A uniform prior in the range $(0, \pi)$ was used for the frequency ω. This is the entire range for the frequency assuming the data were equally sampled. Simulation result

of 5000 samples is displayed in Figure 7.2. The posterior distribution is plotted on top of the histogram. As can be seen, the variance of the posterior distribution is very small, e.g. around 0.004.

Samples of the variance v and the amplitudes a, b are drawn from the corresponding conditional distributions described in Section 7.4. The marginal distribution of a, b are plotted on top of the histogram. Notice that the variance estimate is very big (around 7000), indicating the model does not fit the data very well.

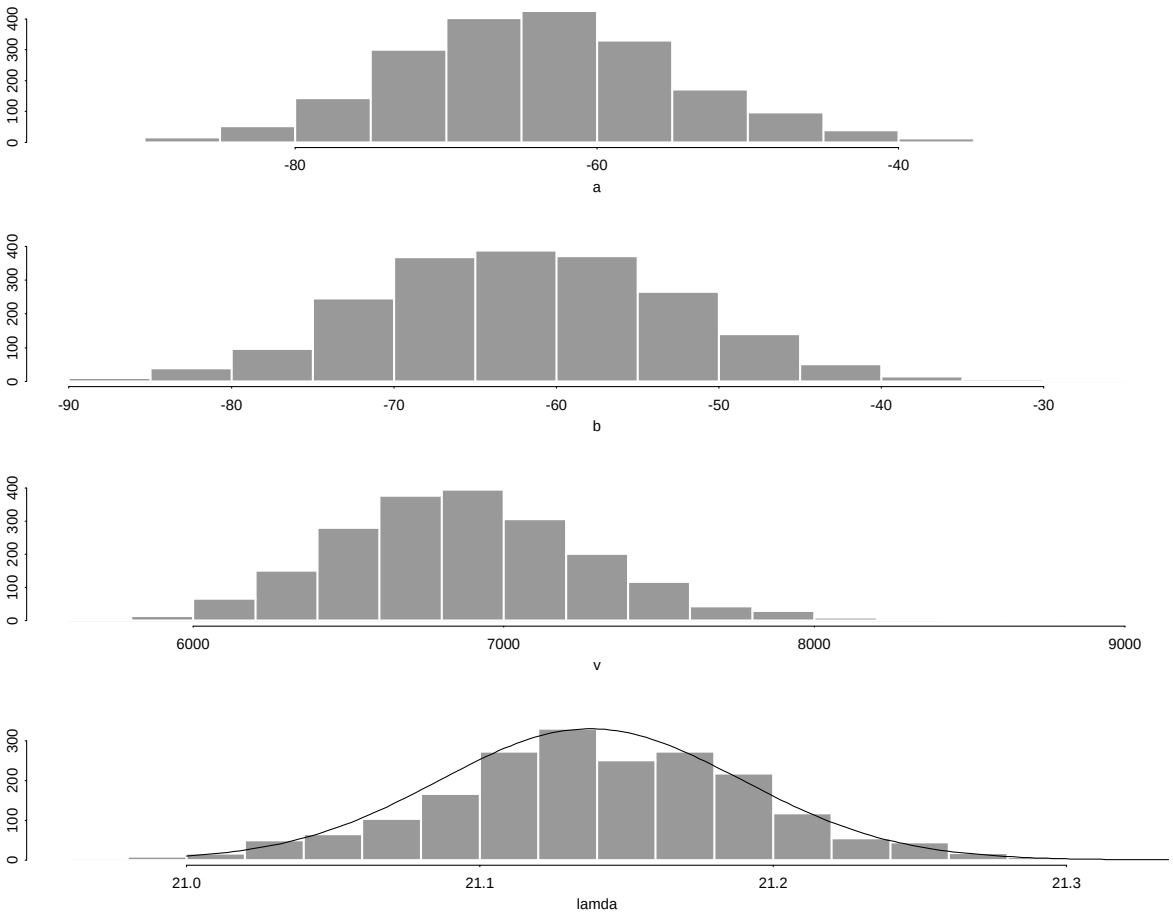

Figure 7.2: *Histogram of parameter estimates for a,b,v and* λ*. With* $p(\lambda|Y)$ *on top of the histogram.*

The residual plot as well as the original data are displayed in Figure 7.3. An obvious cyclical structure of the residual plot indicates the need for further search of

frequencies or further development of the model.

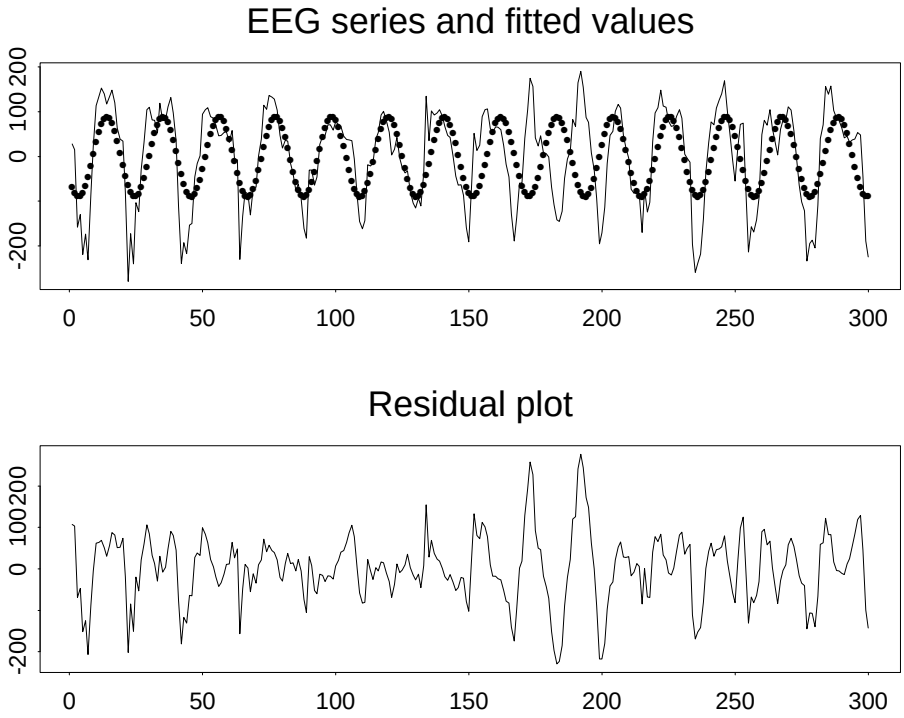

Figure 7.3: *Top: EEG series and fitted values. Bottom: residuals*

A more careful study of the residual plot shows that the cyclical structure is more obvious in the first 100 or so part of the data than in the remaining segment of the 300 data, this indicates the existence of time varying frequency in the data. Plots of $p(\omega|Y)$ in Figure 7.4 for the first 100, last 200, and all 300 of the data shows that the frequencies are indeed varying through time. In particular, the relatively smaller power in the last 200 data suggests a less prominent cyclical structure in the last 200 data.

To find all important frequencies in the data, the same procedure is repeated to analyze the residuals. The posterior distribution of the frequency for the residuals clearly indicates a high power at $\lambda = 17.65$, although the power scale (10^{15}) is much smaller than that of the main frequency (10^{30}). The amplitude for this frequency are

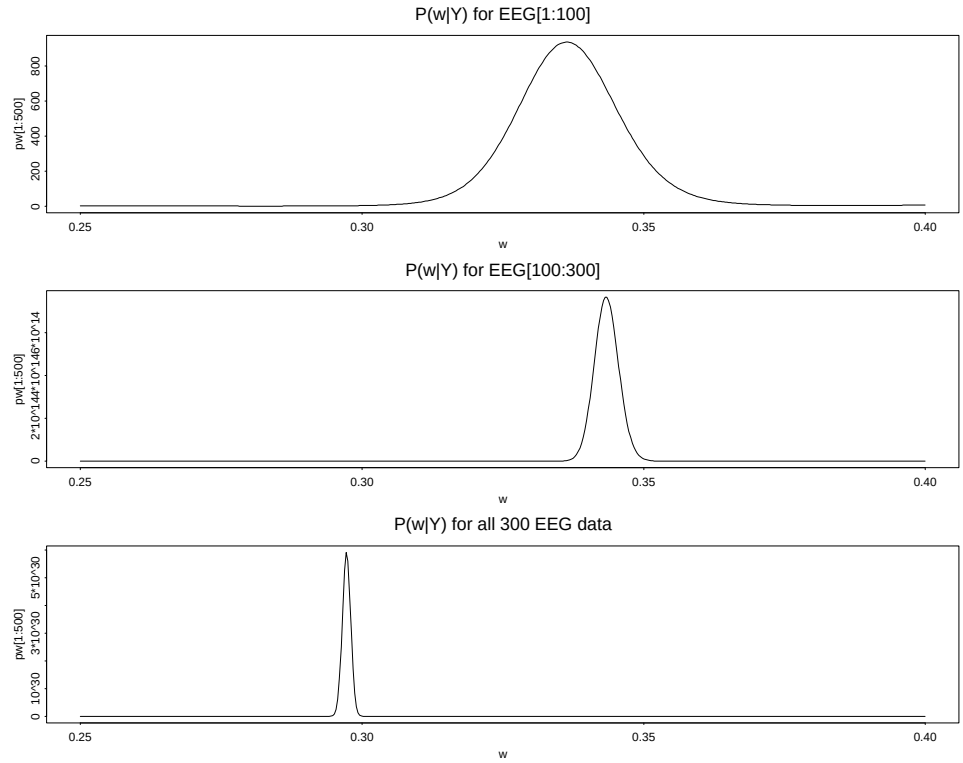

Figure 7.4: *$p(\omega|Y)$ from different segment of the data indicates time-varying frequencies*

about half of that of the main frequency, e.g., around $a = 35$ and $b = -36.9$ and variance estimate for the model reduces to 5600.

An analysis of the residuals for the new model reveals further cyclical structures in the residuals. A new frequency near $\lambda = 19.5$ is singled out, it has almost the same power as the previous frequency (10^{15}) and the amplitude estimates $a = 37.7$ and $b = -33.2$ are similar to those of the second frequency as well. The variance estimate of the model reduces to about 4500.

This procedure was repeated 8 times until no significant reduction in the variance estimate of the model is found. After the 8 frequencies are removed from the data, the acf of the residuals indicates slight lag 1 autocorrelation. See Figure 7.5. The variance estimate of the model after 8 frequencies were removed from the data reduced to about 2500, which is just a third of the original variance estimate. Table 7.1 summarize the

posteriors of the 8 wavelengths.

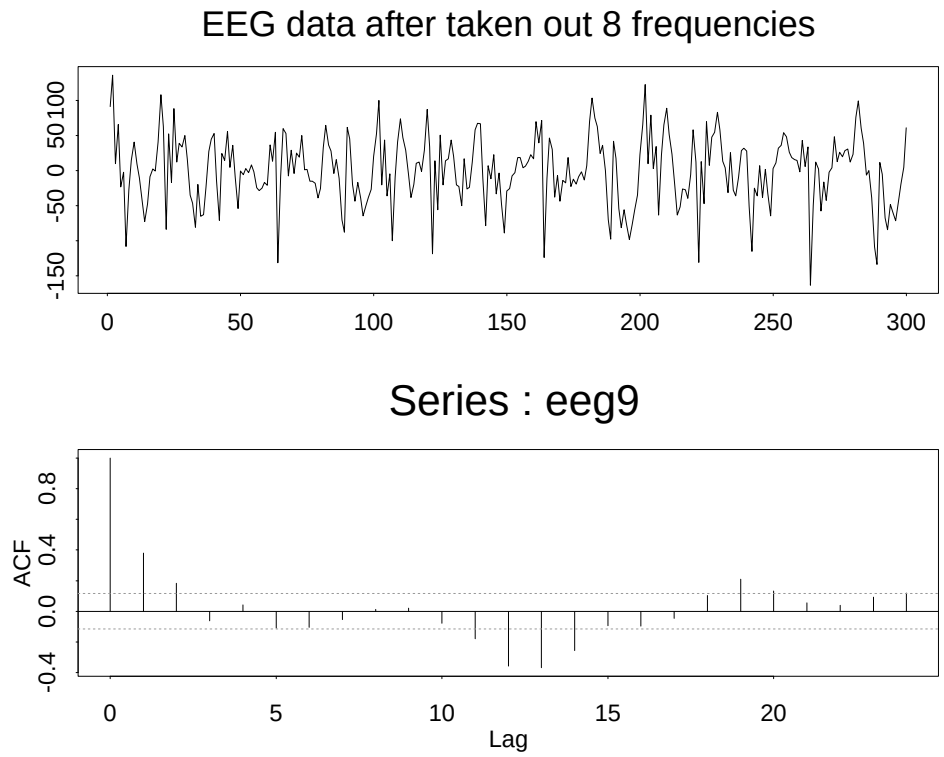

Figure 7.5: *Top: Residual plot of the EEG data after taking out 8 frequencies. Bottom: acf plot of the residuals*

Before any inference can be drawn for the parameters, we check the convergence of the sampler. Since the sampling of the frequency depends only on the data, not on the other parameters a, b, v, each draw of $\{a, b, v, \omega\}$ is independent of the previous one. So the sampling of $\{a, b, v, \omega\}$ is not a Markov Chain. Therefore, there is no need to check the convergence of the entire sampling procedure. However, the sampling of the frequency is based on an Independence Chain Metropolis Sampling, so a convergence check for the sampling of ω is performed.

We apply various convergence check methods to analyze the results using CODA. Since the proposal distribution for the Metropolis Algorithm is very close to the posterior distribution, we expect quick convergence of the sample chain.

A single chain with 10,000 samples was drawn from the the posterior distribution

Table 7.1: Statistics from the predictive residual diagnostic

Wavelength	a	b	V
23.38	-21.35	14.44	3200
21.137	-65.452	-62.72	7000
19.458	37.845	-28.55	4500
17.647	39.467	-38.94	5600
16.18	9.856	-29.56	3900
12.923	23.83	-11.66	2500
11.178	-26.74	0.56	3500
9.657	23.94	-0.15	2800
0.65	3.13	27.56	2.13

of ω using the Independence Chain Metropolis Algorithm. The starting value was drawn from the over-dispersed t-distribution. The summary statistics shows an empirical mean of 21.1 and a standard error at about 0.02. The lag 1 autocorrelation between the batch for every 25 samples is close to zero (-0.0296). This indicates approximate independence between batches is achieved. The autocorrelation between each adjacent sample is only -0.0114, the autocorrelation between every 50 draws can be considered very close to zero (-0.00136). These results support our expectation that the convergence should be very quick for our Metropolis algorithm.

The Geweke convergence diagnostic calculates the mean for the first 10 percent and the last 50 percent of the sample in the chain, then compute the Z diagnostic which is the difference of these 2 means divided by the asymptotic standard error of the difference. A $Z = .0019$ for this chain means that the sampling distribution of Z is very close to a standard Normal distribution. Therefore the chain converged surely. A plot of the Geweke convergence diagnostic shows that almost all Z-scores fall into the 95% confidence interval.

The Raftery and Lewis Convergence diagnostic proposed by Raftery and Lewis (1992) gives the total number of iterations N needed for this chain based on the binomial variance is 3680. The number of initial runs to be discarded as the "burn-

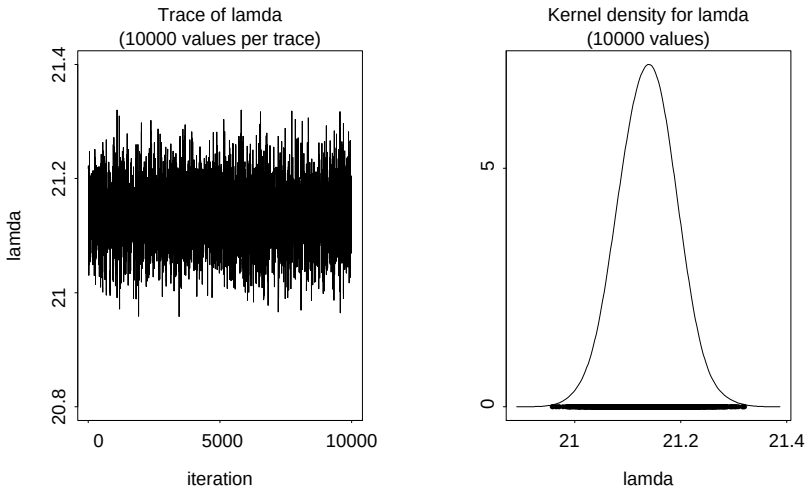

Figure 7.6: *Graphical summary of the output produced by CODA plots*

in" is 2 and the minimum number of iterations needed to estimate the specified quantile to the desired precision if the samples in the chain were independent Nmin is 3746. Here $N < Nmin$, this indicates negative correlation between consecutive iterates which is consistent with the negative correlation founded in the summary statistics. So, 5000 samples used in the previous simulation study can be considered enough.

Finally, another convergence diagnostic, the coupled samples paths proposed by Johnson (1995) is performed. The coupled sampling approach used coupled chains from the same sampler, obtained by using the same sequence of random deviates for each run. The distribution of the iteration at which all sample paths couple gives the appropriate burn-in period needed. Here, 100 starting points were generated from the over-dispersed t-distribution. For the interval between $(0, \pi)$, 100 starting points is believed to be considerably large. The histogram of the number of iterations at which all sample paths couple shows that most chains coupled at iteration 2 which is consistent with the Raftery and Lewis Convergence diagnostic. The 95% quantile of

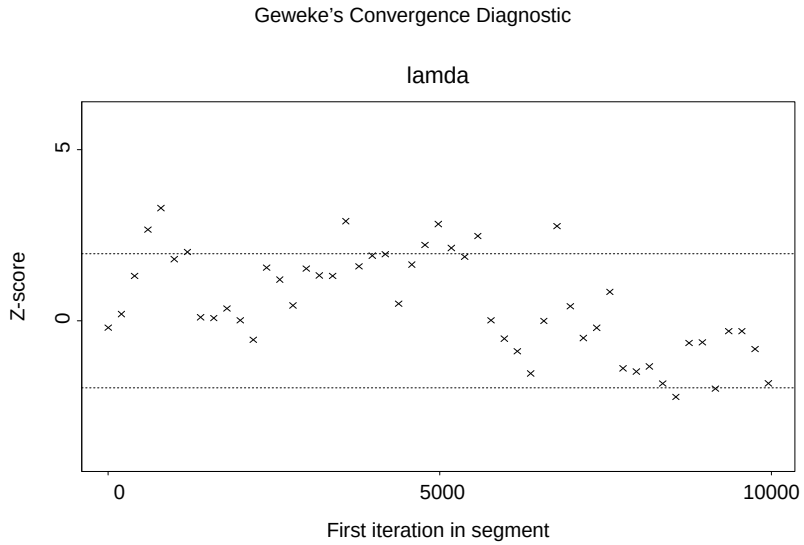

Figure 7.7: *Geweke Convergence Diagnostic*

the number of iterations needed to couple gives 27. There are a few extremely large numbers in the histogram which are over 40, this indicates that the 100 starting values is considerably large.

All these convergence diagnostic shows that the samples drawn from a independence chain Metropolis Algorithm with a very accurate yet over-dispersed candidate generating distribution converge very fast to the stationary distribution. Far Less than 100 draws are needed as the burn-in period.

After the burn-in period, draws from the sampler are considered from the stationary distribution. The summary statistic from CODA shows that the autocorrelation between each adjacent sample is only -0.0114. To be very conservative, we can use samples from every 50th iteration, which has a autocorrelation very close to zero (-0.00136). These draws can be considered independent draws from the posterior distribution and inference of the frequency can be drawn from these samples.

Chapter 8

A Multi-frequency Harmonic Model

In the previous chapter, we discussed the advantage of using Bayesian posterior distribution over traditional Fourier analysis for the estimation of frequencies. We proposed a Gibbs Sampling procedure for the estimation of frequency under the single harmonic model. In this chapter, we extend our analysis to the Multi-frequency Harmonic Model. We develop three sampling procedures to estimate a Multi-frequency model. We then apply the model to the EEG data and Oxygen data.

8.1 The Model

The motivation for the development of multi-frequency harmonic model is obvious. The main advantage of statistical modeling over Fourier analysis is frequency resolution, i.e., the ability to distinguish the spectral response of two or more signals. Traditional Fourier analysis assumes only one frequency in the data, when multiple frequencies especially multiple close frequencies are presented in the data, Fourier analysis may give misleading results. A multi-frequency harmonic model, when properly specified, will identify the multiple signals presented in the data.

A harmonic model with multiple frequencies is

$$y_t = \sum_{j=1}^{J} (a_j cos(\omega_j t) + b_j sin(\omega_j t)) + \epsilon_t \tag{8.1}$$

where $\epsilon_t \sim N(0, v)$. The number of frequencies to be included in the model can

be determined *a priori* from external information or by preliminary checking of data. The frequencies can be well separated or they can be close.

8.2 Parameter Estimation

We illustrate our idea with a two-frequency model, although the methods apply for models with higher number of frequencies.

The parameters to be estimated in model (8.1) are the frequencies $\{\omega_1, \omega_2\}$, the amplitudes and phases $\{a_1, b_1, a_2, b_2, v\}$.

Again, rewrite (8.1) as a linear model

$$Y = X\beta + \epsilon$$

Where $X_{n*4} = (cos(\omega_1 t), sin(\omega_1 t), cos(\omega_2 t), sin(\omega_2 t))$, $\beta = (a_1, b_1, a_2, b_2)'$ and ϵ is the diagonal variance-covariance matrix with variance v as each element on the diagonal.

Given reference prior $p(\beta, \sigma^2) \sim 1/\sigma^2$, the posterior distribution for the frequencies under a uniform prior for the frequencies is

$$p(\omega_1, \omega_2|Y) \quad \propto \quad |X'X|^{-1/2}(1 - (\hat{\beta}'X'X\hat{\beta})/(Y'Y))^{(4-n)/4} \tag{8.2}$$

The structure of (8.2) is similar to (7.4) in the single-frequency model. It is highly concentrated at major frequencies and has numerous local peaks. No side lobes are presented in the posterior distribution around the frequencies because the data was not windowed. When there are two frequencies in the data, the contour of joint posterior distribution has two symmetric "eyes" at the two frequencies for both well-separated or close frequencies.

A t-distribution approximation was again used by Bretthorst (1988) to perform analytical inference of the frequencies. The approximation works well when the two frequencies are well separated, but no solution was given for short data set with two close frequencies, the case when traditional Fourier transforms fail to work. To achieve accurate estimation of the frequencies, sampling-based inference has to be developed.

Sampling from a multi-modal high dimensional joint posterior distribution is not an easy task. The grid method to search for modes developed in the previous chapter is not appropriate for high dimensional distributions. We propose three Gibbs samplers to sample the joint posterior distribution. Let $\theta = \{a_1, b_1, a_2, b_2, v\}$, these methods are:

Gibbs Sampler 1: Sample joint distribution of the frequencies conditioning only on the data

- sample $p(\omega_1, \omega_2|Y)$
- Sample $p(\theta|Y, \omega_1, \omega_2)$

Gibbs Sampler 2: Sample the conditional distribution of the frequency conditioning only on the data

- $p(\omega_1|\omega_2, Y)$
- $p(\omega_2|\omega_1, Y)$
- Sample $p(\theta|Y, \omega_1, \omega_2)$

Gibbs Sampler 3: Sample the frequencies conditional on the data and model parameters

- $p(\omega_1|\omega_2, a_2, b_2, Y)$
- $p(\omega_2|\omega_1, a_1, b_1, Y)$
- Sample $p(\theta|Y, \omega_1, \omega_2)$

We discuss each of these in the following sections.

8.2.1 Sampling of the Joint Posterior

When there are only two frequencies in the model, we can sample the joint distribution directly. The marginal posterior $\{\omega_1, \omega_2\}$ is then

$$p(\omega_1, \omega_2|Y) \quad \sim \quad |X'X|^{-1/2}(1 - (\hat{\beta}'X'X\hat{\beta})/(Y'Y))^{(4-n)/4} \tag{8.3}$$

The shape of the joint distribution has high conical peaks at the frequencies and has numerous local conical peaks elsewhere. It might have two symmetric peaks at the two most important peaks if there are only two frequencies or it might have more peaks if the data has multiple frequencies. We need a sampling algorithm to efficiently traverse the whole probability surface and sample all the modes.

As stated earlier, (8.3) is often highly concentrated within a very small or several very small intervals, and has numerous small local modes everywhere. Such character makes it easy for the standard random walk Metropolis algorithm to stay at a local mode. Therefore, standard random walk Metropolis algorithm will not be efficient since a candidate generating distribution with large variance will result in high rejection rate at local modes, and one with small variance will not be able to traverse the whole surface efficiently. Instead, the Independence Chain Metropolis algorithm is again used.

Because of the exponential shape of (8.3) at each mode, a mixture of t-distribution centered at the modes of (8.3) is selected as the candidate generating distribution will can best mimic (8.3). For a joint distribution with higher dimension, direct mode-finding of the joint distribution isn't an easy task though.

The gridding method was used in the last chapter to search for modes. It is reasonably fast in the one dimensional case especially when it has to be performed only once. In higher dimensional case, the grid size increase exponentially and the

computation becomes impractical. The drawback of this method is that it spends an even amount of effort on the huge unimportant areas and the very small important areas where the frequencies lie.

To search for modes efficiently in the entire probability surface, we need a method which traverse the whole surface yet can move quickly to the modes. We propose a random walk Metropolis-like algorithm which serves this purpose. We borrow the idea of the Metropolis algorithm as a method to traverse the probability surface. Yet we make three modifications to the algorithm to speed up the movement of the algorithm for the purpose of mode-finding rather than sampling the distribution.

The grid method evenly divides the probability surface and search for modes. When evaluation of the distribution is expensive, this slows down the process considerably. In contrast, the random walk Metropolis algorithm moves towards the modes of the distribution by assigning higher acceptance probability to samples. The dilemma with the random walk Metropolis algorithm is the coverage of the probability surface versus rejection rates. Namely, a random walk Metropolis algorithm with a large variance for the proposal distribution will ensure coverage of the probability surface, yet the rejection rates will be high when a sample near the modes is found.

Since the purpose here is to search for modes rather than sampling the distribution, the dilemma with the random walk Metropolis is solved. We modify the algorithm to only accept samples with higher density values. This ensures the sampler moves only towards the direction where a mode lies.

Another problem with the random walk Metropolis algorithm is that it is easy for the random walk Metropolis algorithm to stay at local modes when the distribution is multi-modal. Again, for the purpose of mode searching, this can be avoided by finding modes sequentially rather than simultaneously. This means that after a local mode is found, the corresponding frequency can be removed from the data. The same

procedure can then be applied to the posterior distribution of the frequency of the new data.

A detailed description of the method is as follows:

- Generate one sample y_{n+1} from a proposal distribution centered at the current draw y_n with a reasonably large variance.
- Evaluate the value of the posterior distribution at both the current draw y_n and the new sample y_{n+1}, calculate the ratio: $p(y_{n+1})/p(y_n)$, accept y_{n+1} if the ratio is bigger than 1.
- If the ratio is significantly big, it means a sample near a mode is found. The sample can be used as a starting value for the Newton's method to find the exact mode of the distribution.
- Save the mode found, then remove the corresponding frequency from the data. And the above procedure can be performed again to find other modes.

This method of searching for modes is useful especially for high dimensional posterior distributions when the computation for the grid size method increases exponentially.

Since the joint posterior distribution of ω_1, ω_2 depends only on the data, the joint distribution (8.2) doesn't change in each iteration in the Gibbs Sampler, the above routine for finding modes needs to be performed only once before the Gibbs Sampler starts. The mixture of t-distribution resulted from the method will mimic the target distribution very well. Therefore, the convergence of the independence Chain Metropolis sampler will be very fast.

8.2.2 Sample $\{\omega_1, \omega_2\}$ iteratively

An alternative approach to sample the joint distribution that can be generalized to higher dimensions is to sample $p(\omega_1, \omega_2, |Y)$ indirectly by iterating through the conditional posterior distributions

- $p(\omega_1|\omega_2, Y)$
- $p(\omega_2|\omega_1, Y)$

Without loss of generality, assume $\omega_1 > \omega_2$. Given an uniform prior, the conditional posterior distribution can be found from the joint distribution:

$$\begin{aligned} p(\omega_1|\omega_2, Y) &\sim |X'X|^{-1/2}(1 - (\hat{\beta}'X'X\hat{\beta})/(Y'Y))^{(4-n)/2}, \qquad \omega_1 > \omega_2, \\ &= 0, \qquad \omega_1 < \omega_2. \end{aligned}$$

And similarly,

$$\begin{aligned} p(\omega_2|\omega_1, Y) &\sim |X'X|^{-1/2}(1 - (\hat{\beta}'X'X\hat{\beta})/(Y'Y))^{(4-n)/2}, \qquad \omega_2 > \omega_1, \\ &= 0, \qquad \omega_2 < \omega_1. \end{aligned}$$

Here $p(\omega_1|\omega_2, Y)$ and $p(\omega_2|\omega_1, Y)$ have the same structure as the marginal posterior $p(\omega|Y)$, with an additional constraint of the order of the two frequencies. Therefore, the same sampling methods can be used to sample from $p(\omega_1|\omega_2, Y)$ and $p(\omega_2|\omega_1, Y)$.

In Chapter 7, sampling of the posterior distribution of the frequency given the data is based on the Independence Chain Metropolis Algorithm using a mixture of t-distributions as the candidate generating distribution. Various methods were proposed to find the modes of the candidate generating distribution. Since the conditional posterior distribution of each frequency depends only on the data and the

other frequency, the mode finding procedure needs to be done only once with slight modification according to the newly sampled frequency.

8.2.3 Sample $\{\omega_1, \omega_2\}$ conditioning on other parameters

A slight modification to Gibbs Sampler 2 is to allow the the frequencies conditioning on the data as well as on the amplitudes and phases. The setup of the Gibbs Sampler 3 is as follows:

- $p(\omega_1|\omega_2, a_2, b_2, Y)$
- $p(\omega_2|\omega_1, a_1, b_1, Y)$
- $p(1/v|Y, \omega_1, \omega_2) \sim Ga((n-4)/2, ns^2/2)$
- $p(a_1, b_1, a_2, b_2|Y, \omega_1, \omega_2, v) \sim N(\hat{\beta}, (X'X)^{-1}v)$

Given $\{\omega_2, a_2, b_2\}$, the corresponding cosine wave $a_2 cos(\omega_2 t) + b_2 sin(\omega_2 t)$ can be removed from the data. The reminding part $\hat{y}$ should follow the single frequency model

$$\hat{y}_t = a_1 cos(\omega_1 t) + b_1 sin(\omega_1 t) + N(0, v)$$

if the model is correct.

So the conditional posterior distribution of ω_1 given $\{\omega_2, a_2, b_2\}$ is the conditional posterior distribution of ω_1 given $\hat{y}$. Again, this is a one-frequency model and the conditional posterior distribution of ω_1 given $\hat{y}$ is the same as (7.3), i.e.,

$$p(\omega_1|\omega_2, a_2, b_2, Y) \sim |X'X|^{-1/2}(1 - (\hat{\beta}'X'X\hat{\beta})/(\hat{Y}'\hat{Y}))^{(2-n)/2}. \quad (8.4)$$

Therefore, the same sampling methods from Chapter 7 can be used to sample (8.4). Since each iteration of the Gibbs Sampler uses different data, the Griddy method needs to be repeated for each iteration of the Gibbs Sampler.

In comparison with Gibbs Sampler 2, removing the given cosine wave from the data and sample from the new posterior distribution from the new data is more intuitive as we "filter" out cosine waves sequentially. On the other hand, sampling the frequencies given only the data theoretically converge faster to the joint posterior distribution since the other parameters have been integrated out. But since the functional form of both distributions and the methods used to sample them are the same, the convergence rate is similar. Simulation experience supports this argument.

8.3 Application

We now apply the multi-frequency harmonic model to analyze some cyclical data. We first analyze the EEG data used in the previous chapter. Then we apply the model to some Oxygen data in geological studies.

8.3.1 EEG Data

In the previous chapter, we applied the single-frequency harmonic model to the EEG data repeatedly to find 8 frequencies. We also found evidence of time-varying wavelengths. This suggests the existence of multiple frequencies in the EEG data. Of the eight frequencies, three or four frequencies are probably important. This preliminary analysis can be used to help determine the starting point of the frequencies and amplitudes in the multiple frequencies sampling procedure and to help identify the number of frequencies in the data.

We now apply the multi-frequency harmonic model to the EEG data. As an illustration, we start with a two-frequency model. As outlined in section 2, there are multiple ways to sample the joint posterior distribution. We adopted the second approach, which samples the two conditionals $p(\omega_1|\omega_2, Y)$ and $P(\omega_2|\omega_1, Y)$. This approach doesn't require starting values for a_1, b_1, a_2, b_2.

Since the posterior distribution is symmetric about ω_1 and ω_2. We assume $\omega_1 < \omega_2$ and consider only one side of the joint posterior distribution.

The independence chain Metropolis sampling algorithm was used to sample the conditional posterior distribution $p(\omega_1|\omega_2, Y)$. The mode finding procedure was constructed with the range constrained by ω_2. And the same procedure was performed for $p(\omega_2|\omega_1, Y)$.

5000 samples of λ_1, λ_2 were drawn. Figure 8.1 shows the histograms of the model parameters. The two most important wavelength $\lambda = 21.137$ and $\lambda = 19.354$ are found. The corresponding amplitudes of the frequencies are around $a_1 = 35, b_1 = -30$ for $\lambda = 19.354$. And $a_2 = -62.5$, $b_2 = -55$ for $\lambda = 21.137$. The variance estimate of the model is centered at $v = 6000$. About 1000 less than that of the single-frequency model.

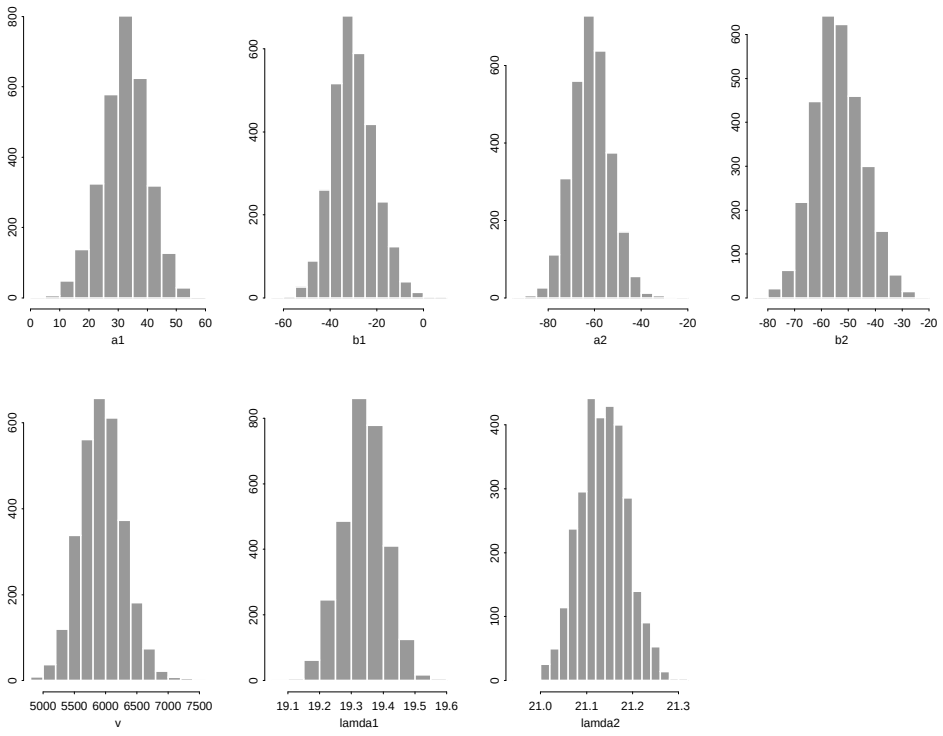

Figure 8.1: *Histogram of model parameters for two-frequency model on EEG data*

Next, we apply a four-frequency model to the data. We apply a Gibbs Sampler to sample $p(\omega_1, \omega_2, \omega_3, \omega_4|Y)$ by iterating through

$P(\omega_1|\omega_2, \omega_3, \omega_4, Y)$

$P(\omega_2|\omega_1, \omega_3, \omega_4, Y)$

$P(\omega_3|\omega_1, \omega_2, \omega_4, Y)$

$P(\omega_4|\omega_1, \omega_2, \omega_3, Y)$

conditional on $\omega_1 < \omega_2 < \omega_3 < \omega_4$.

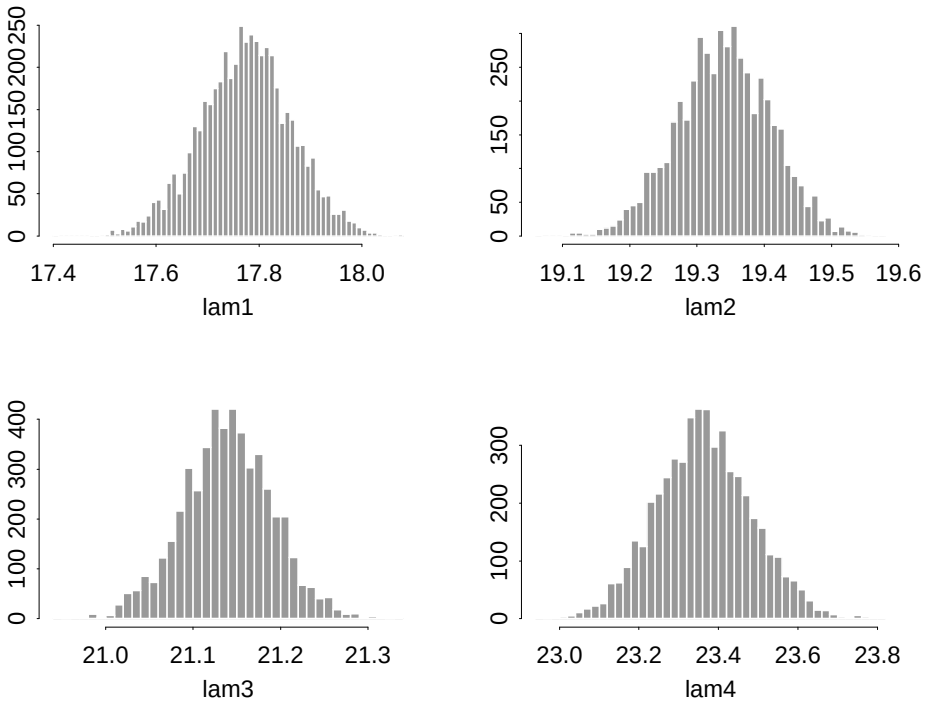

Figure 8.2: *Histogram of the four wavelengths of the EEG data*

5000 samples of $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ were drawn. These results were then used to sample the variance and the amplitudes. The histogram of the four wavelengths and the amplitudes are in Figure 8.2 and 8.3. The four most important frequencies are near $\lambda_1 = 21.137$, $\lambda_2 = 19.354$, $\lambda_3 = 17.774$ and $\lambda_4 = 23.4$. The corresponding amplitudes of the frequencies are around $a_1 = 8.6$, $b_1 = -57.2$ for $\lambda = 17.774$,

$a_2 = 42.5$, $b_2 = -25$ for $\lambda = 19.354$, $a_3 = -62.5$ $b_2 = -55$ for $\lambda = 21.137$ and $a_4 = -30$,$b_4 = 15$. The variance of the model is centered at $v = 4000$. Far less than the 7000 from the single-frequency model. The rejection rates recorded for the Metropolis algorithm for each λ_j, $j = 1..4$ were $30\%, 19\%, 24\%, 33\%$.

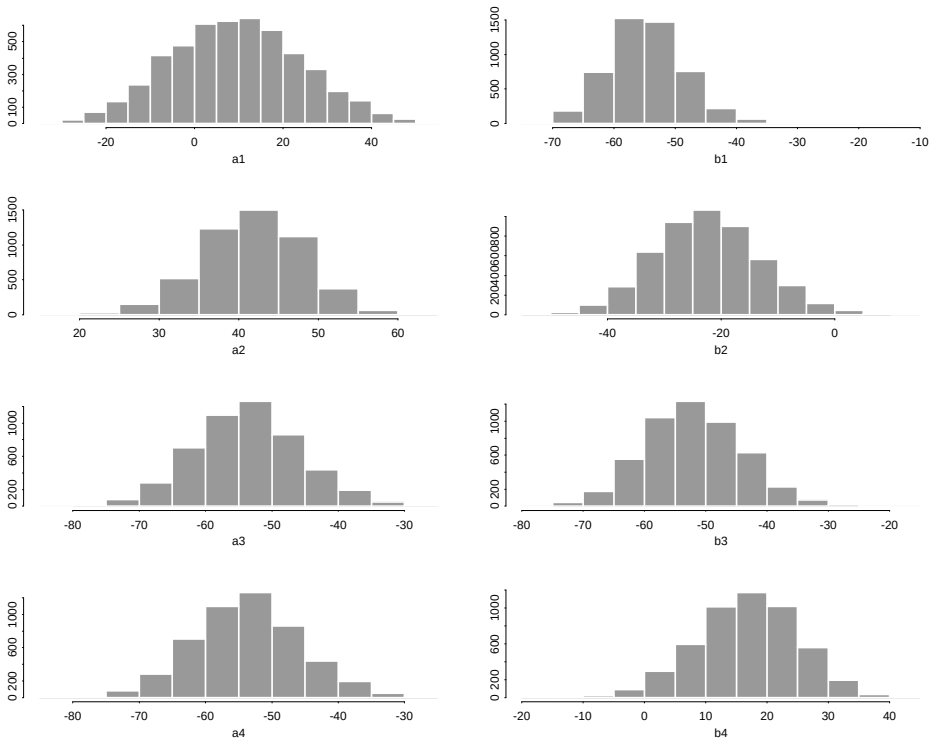

Figure 8.3: *Histogram of the amplitudes for the four-frequency model*

8.3.2 Oxygen Data

Next, we apply the multi-frequency model to some Oxygen data studied by West (1996). Geological time variations in oxygen and other isotope measurements from deep ocean cores relate to patterns of variation in global ice volume and ocean temperature (See Shackleton and Hall 1989, Park and Maasch 1993). Periodic and quasi-periodic behaviors are usually observed in such series. We study a time series of single deep ocean core oxygen isotope which is representative of several oxygen isotope se-

ries from cores of various geographical locations. See Figure 8.4 for a time series plot of the data with the mean subtracted. This data series is derived from original $\delta^{18}O$ site 677 measurements presented in Shackleton and Hall (1989) and was provided by Park of Yale University. The values estimate relative abundance of $\delta^{18}O$ and are timed on an equally spaced $3k$ year scale. The time scale is based on that of Ruddiman, McIntyre and Raymo (1989), and is discussed in Park (1992). Park and Maasch (1993) discusses the process of interpolation of original, unequally spaced measurements to this equally spaced scale. This series dates back to roughly 2.5 million years, and is plotted in reverse of sign, by convention, so that the apparent increase in level in model times reflects generally warmer average global temperatures and smaller average ice masses. There are surely errors in the timing of the observations, due to the process of estimating true calendar times of the geochemical data as well as calibration times to the nearest unit. (3000 years here). West (1996) applied a state-space autoregressive model which allows time-varying amplitudes and multiple periodicities to study the data. Here, we assume the imputed equally spaced times are accurate and study the series using a multi-frequency harmonic model in the frequency domain. In the next chapter, we study the impact of timing error in the estimation of frequencies. Since time-varying amplitudes are obvious, we divide the data into two parts, and study the first 433 data.

Plots of the log-posterior distribution for the wavelength for the first 433 Oxygen data in Figure 8.5 shows two dominant wavelengths near $\lambda = 95$ and $\lambda = 41$. With the one near $\lambda = 95$ being more important. This period around 100 kyears (also called the "100,000-year ice-age cycle") is of major interest and has been the subject of intensive investigation in recent years. Identifying the nature and structure of quasi-periodic components of period around 100 kyears is of importance in contributing to debates over the genesis of the ice-age cycles, roughly a million ago, and to questions

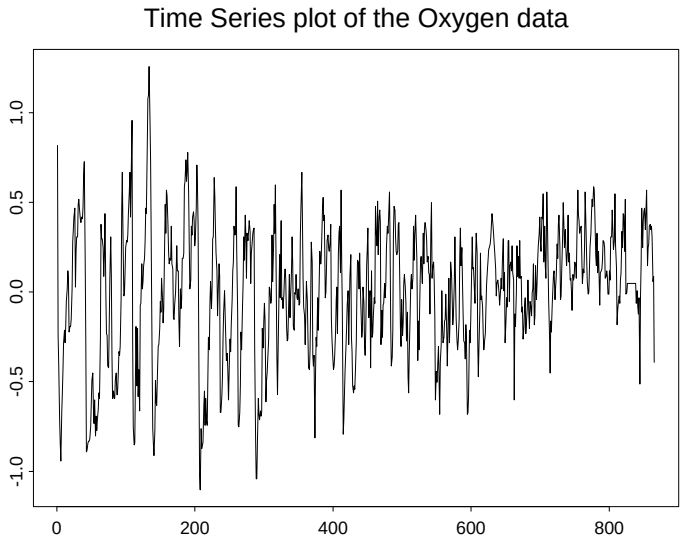

Figure 8.4: *Time series plot of the Oxygen data*

of whether or not the onset was gradual and inherent or the result of a significant structual climatic change. (See Ruddiman et al 1989 and Park 1992 for discussion on this).

The log-posterior distribution indicates other small wavelengths at around $\lambda = 120$ and $\lambda = 23$.

we first apply a two-frequency harmonic model to the data since two prominent frequencies appear in $p(\omega|Y)$. A Gibbs Sampler which iterates through $p(\omega_1|\omega_2, Y)$ and $p(\omega_2|\omega_1, Y)$ was run for 3000 iterations after discarding the first 500 iterations as the "burn-in" period. Figure 8.6 shows the posterior distribution of the two wavelengths centered at $\lambda_1 = 41.2$ and $\lambda_2 = 95.5$. The corresponding amplitudes are $a_1 = 0.05$, $b_1 = -0.22$ for λ_1 and $a_1 = 0.19$, $b_1 = 0.15$ for λ_2. The variance is around $v = 0.14$. The autocorrelation for every 10 draws is very small (-0.025 for λ_1 and -0.004 for λ_2). The Z scores for Geweke's convergence check were very close to zero, indicated that 3000 draws were enough for the simulation. The Raftery and Lewis convergence diagnostic gave the burn-in period 5 and 9 for λ_1 and λ_2.

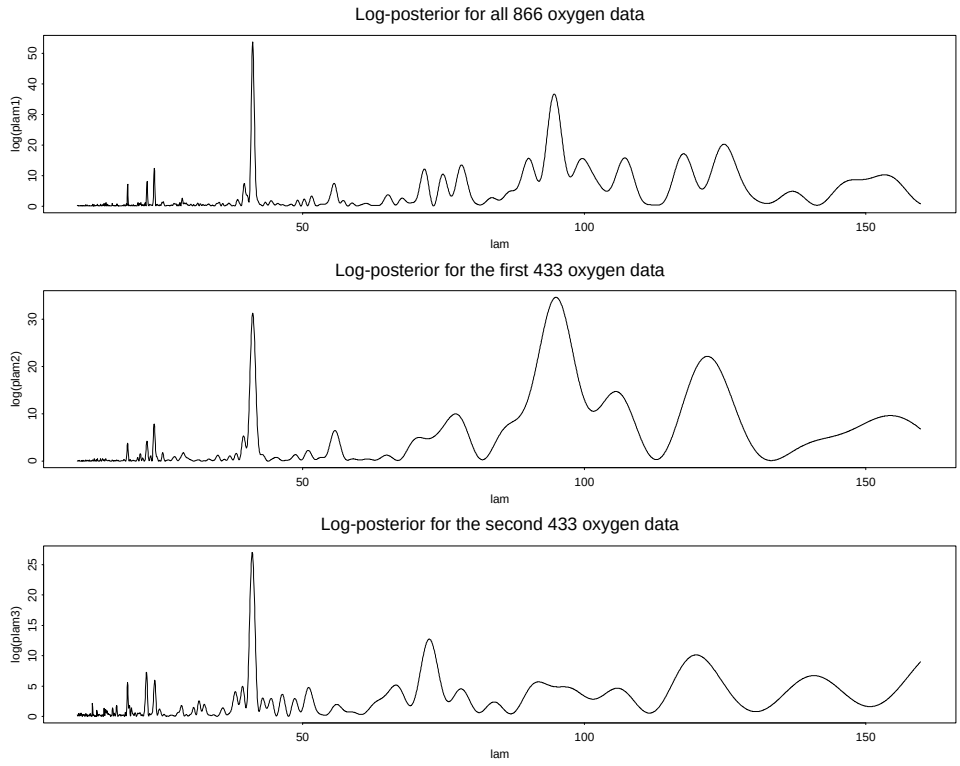

Figure 8.5: *Posterior $p(\lambda|Y)$ for all 866 Oxygen data and for the first and second half of the Oxygen data*

The rejection rates for the Metropolis algorithm are 38% and 28% for λ_1 and λ_2 respectively.

Next, we apply a four-frequency harmonic model to the data. 2000 draws from the Gibbs Sampler were used after discarding the initial 500 samples. Figure 8.7 shows the histogram of the posterior distribution for the four wavelengths. They are centered at $\lambda_1 = 23.6$, $\lambda_2 = 41.2$, $\lambda_3 = 95.7$ and $\lambda_4 = 121.5$. The variance for each posterior distributions are $v_1 = 0.058$, $v_2 = 0.012$, $v_3 = 0.376$ and $v_4 = 0.59$. A 95% confidence interval for each wavelength is $(23.49, 23.77)$, $(40.89, 41.35)$, $(94.56, 96.96)$ and $(120.01, 123.07)$ respectively.

These results are consistent with the results of West (1996) in the time domain analysis. The shorter periodicies correspond to the precession (19-23k years) and obliquity (40-42k years) of the earth's orbit impact on insolation received by the earth

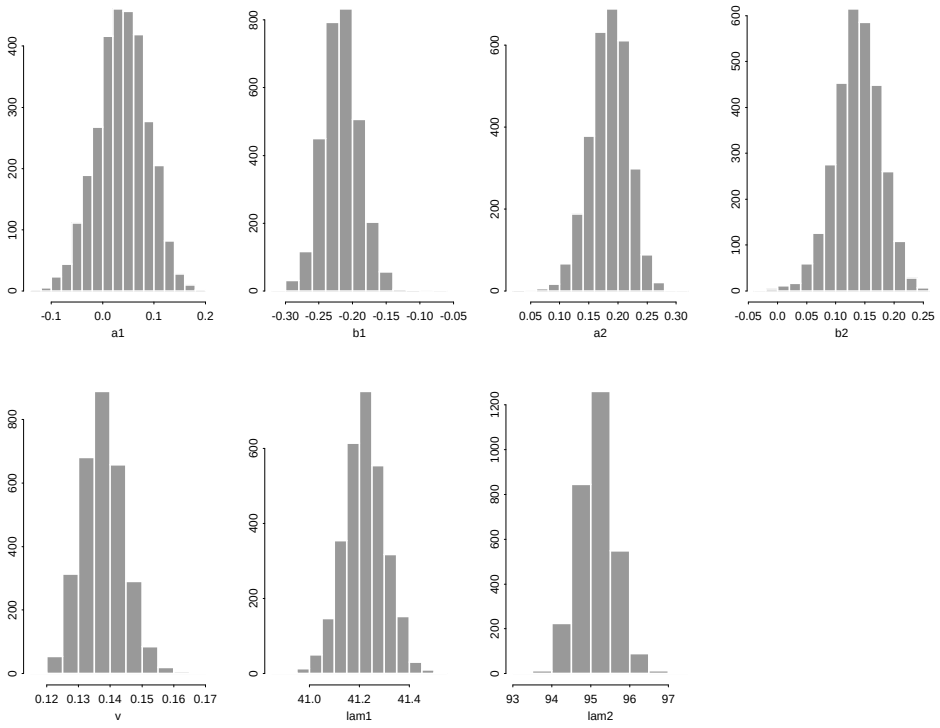

Figure 8.6: *Histogram of model parameters for the two-frequency model of the first half of the Oxygen data*

which induce substantial variations in climate characteristics. The longer periodicities correspond to periodicities in eccentricity associated with periods of around 95-100k years and 120-130k years. Our study supports the view that the 100 kyears ice-age cycle is dominant, followed by cycles related to the obliquity (41 kyears) and precession (23-24 kyears).

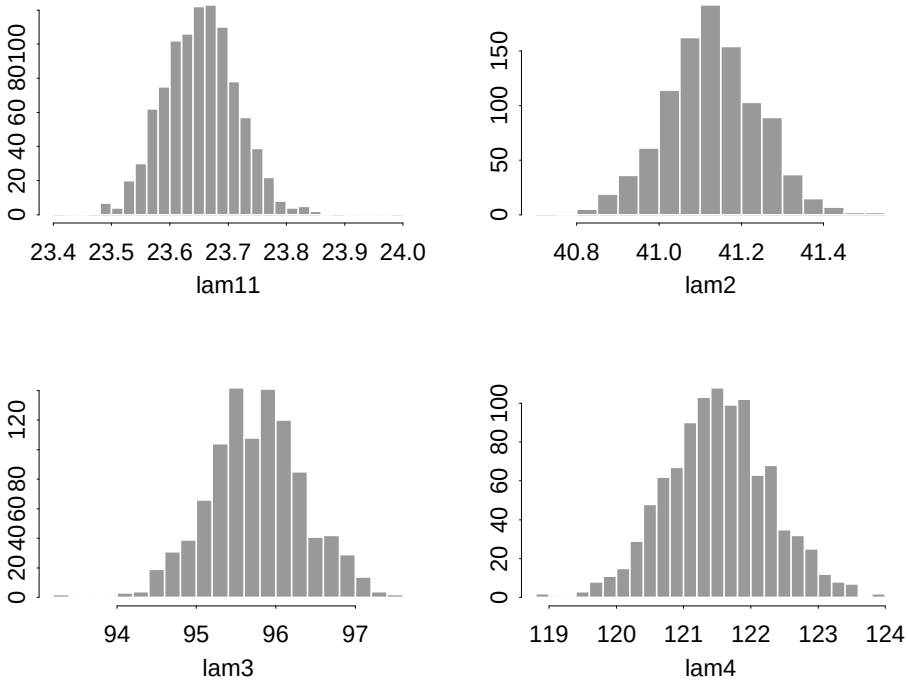

Figure 8.7: *Histogram of the four wavelengths for the four-frequency model of the first half of the Oxygen data*

Chapter 9

Timing Uncertainty

In the previous chapters, we studied the problem of frequency estimation under equally spaced time scale. We analyzed some cyclical data including the EEG data and Oxygen data using our methods. In the application to the Oxygen data, we addressed the issue of uncertain timing in the data. In this chapter, we investigate the impact of uncertain timing to the study of frequency estimation. We develop models that incorporate the uncertain timing and develop methods to estimate frequency and time under the model. This chapter is organized as follows. First, we discuss the problem of timing uncertainty in time series studies. Second, we study the impact of uncertain time on frequency estimation in harmonic model studies. Third, we develop an harmonic model under uncertain timing and develop methods of conducting statistical inference on timing. We then apply this model to the Oxygen and EEG data analyses.

9.1 Timing Issues in Time Series Studies

One important feature of time series modeling is the timing of the data. Traditional time series models usually assume equally spaced observations and the timing is precise. These two assumptions are violated in many applications.

In some cases, there are timing irregularities which occur when observations arrive at intervals of irregular length. For example, monthly data may be recorded on a particular day towards the end of each month, but the precise day can vary

between each months. Daily observations may be recorded only to the nearest hour. In financial time series, data are not observed on weekends and holidays. In these cases, where the data is reported at an aggregated level but may be based on different numbers of time intervals at a lower level of aggregation, it's clear that the timing should be accounted for in the model. One way to deal with unequally spaced data is to identify a base, minimum time interval underlying the observed times of observations and develop the equally spaced model on that time scale. Then the times with no observations are treated as points of missing data. Other ways to deal with unequally spaced data include imputing data, etc.

In other applications, the timing of the data is not precise. This happens when the timing of the data is recorded indirectly through some other measurements. Some of the timing errors are random, while others are systematic. For example, timing uncertainty is rooted in every study of geochemical time series. Mclaren (1976), stated that "Geology has come a long way without an accurate time scale". In geological study, geologists collect samples fossils, rock units and mineral bodies, they infer their position in a geological succession, and use a procedure to correlate the relative position to a relative time scale. There are many error sources in this whole process. The formation of fossils, rocks and mineral bodies may be affected by some irregular climate and geologic change or environmental contaminations which introduces error in the inference of the position or formation of the sample. Also, the response of sedimentary records to some periodic forcing may not be a linear one which may distort the true spectrum of the core. In addition, the various procedures used by geologists to calibrate the time given a relative position of a core including biochronology, radiometry and magnetostratighaphy have built-in errors. For example, it is well known that the rather imprecise procedure involved in depth-to-radiocarbon dating creates significant timing uncertainty for geochemical time series analysis in

general. Also, perfect sampling of core is hardly achieved which gives additional source of uncertainty regarding to the relative position of the samples. In summary, the errors of timing calibration can come from uneven core sampling, uncertainty in the measuring procedure, bioturbation, non-periodic climate and geologic forcing mechanisms, and the nonlinear response of the geologic variables to periodic forcing. See Mclaren (1976) for a discussion of timing in geologic study. All these errors can have significant effect on the precision of the formal statistical analysis of the time series.

West(1995) studied a geological time series of proxy records of variations in climatic conditions. The series is a geochemical quantity, from deep lake sediment, that is a proxy indicator of local climatic conditions of interest in studying climatic change over time. The data were derived from the original $\delta^{18}O$ measurement from Site 677 of the Deep Sea Drilling Project presented in Shackleton and Hall (1989) and was provided by Park of Yale University. The data was then mapped to calendar ages by a complex and uncertain process involving carbon-14 calibration discussed in Ruddiman, McIntyre and Raymo (1989). Notice that the time must have been truncated to the nearest 3 kyear since all the times are integers. Finally, the unequally spaced calendar time series were interpolated to an equally spaced series for analysis by Park and Maasch (1993). The final values estimate relative abundance of $\delta^{18}O$ and are timed on an equally spaced, 3k year scale. Each of the above processes brings some uncertainty into the final timing of the data. A plot of the $\delta^{18}O$ measurement series versus the original depth of the core measurement is plotted in Figure 9.1. Partial plots of $\delta^{18}O$ measurement series versus the unequally spaced time and equally spaced time are in Figure 9.2. The difference of the two series is obvious.

Shackleton and Hall (1989) discussed some of the errors in the sampling of the core including the quality of the core, and the error from the unequal sampling of

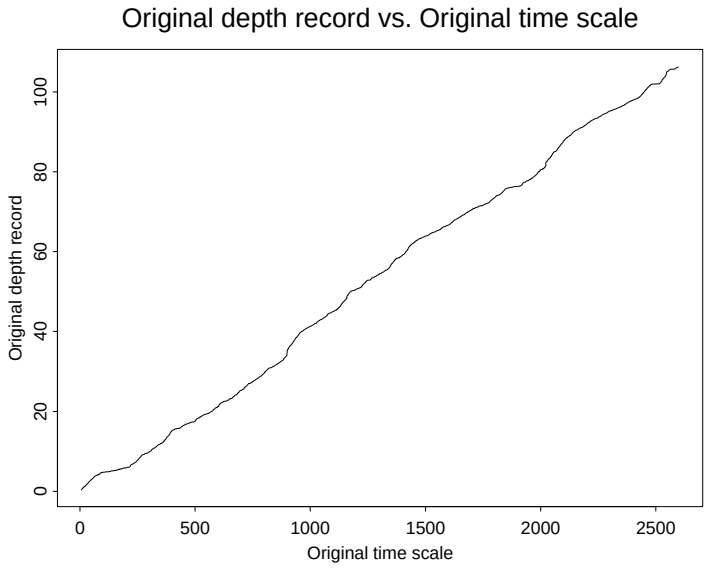

Figure 9.1: *Original depth of the core measurement versus $\delta^{18}O$ measurement series*

the core and gaps between cores which resulted in the uncertainty of the inter-core depth. Other errors come from carbon-14 calibration of the core, and the error from the interpolation of unequal time scale to equal time scale. The magnitude of each error for this data were not recorded, which is a common problem for many geology studies. Harland et al.(1990) pointed out that in many geology studies, "The problem is sometimes figuring our just what the original authors intended. Some don't even report errors, others fail to state explicitly which error they are using, assuming everyone understands exactly what they do." A conservative estimate of the magnitude of the total error is about 10%, though in most actual cases, the error could be much bigger.

9.2 Timing Uncertainty in Harmonic Models

From the above discussion, it is clear that timing for most geological time series are subject to considerable error. Such timing error may have dramatic impact on the accuracy of the study. In particular, in the context of harmonic models, when the

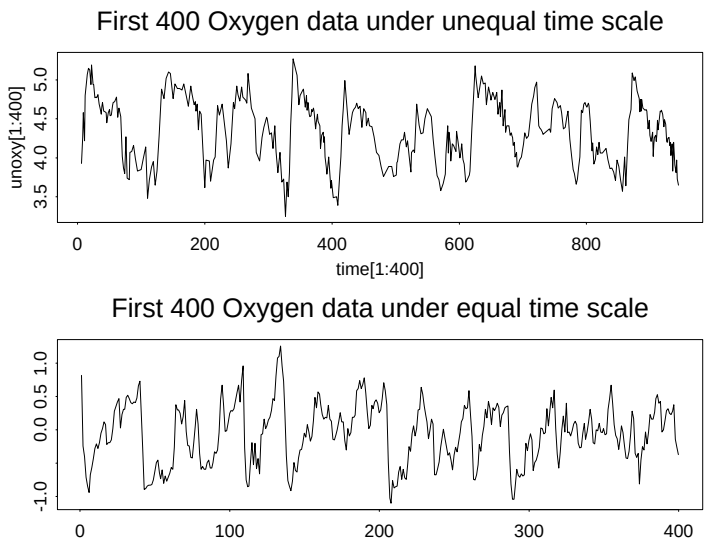

Figure 9.2: *The first 400 oxygen data under unequally spaced time 3y scale vs. under equally spaced time scale*

interest of the series is on the periodic property of the series, timing error may affect the estimation of the wavelength of the series under the harmonic model developed in this thesis. Study on spectral analysis of irregularly spaced data exists, Perzen (1983) provides a collection of papers in this topic. The key paper by Brillinger (1972) lays down a theoretical framework and also suggests a general method for non-parametric spectral analysis in the case where the actual sampling times are known. The practical use of Brillinger's methods are discussed by Moore *et al.* A discussion of the impact of irregularly spaced sampling on spectra in practice and in theory can be found in Moore and Thomson (1991). In this section, we discuss the impact of irregularly spaced sampling on spectra from a Bayesian point of view by studying the posterior distribution of the frequency $p(\omega|Y)$.

Refer to the single-frequency harmonic model in Chapter 7,

$$y_i = acos(ws_i) + bsin(ws_i) + \epsilon_i$$

Here $S = \{s_1, s_2, ...s_n\}$ represents the observed timing of the data. S might be

equally spaced, or unequally spaced, and subject to error. The posterior distribution for the frequency $p(\omega|Y,S)$ depends on the timing S through the observation matrix X. The posterior distribution for the frequency

$$p(\omega|Y) \quad \sim \quad |X'X|^{-1/2}(1-(\hat{\beta}'X'X\hat{\beta})/(Y'Y))^{(2-n)/2} \tag{9.1}$$

depends on S.

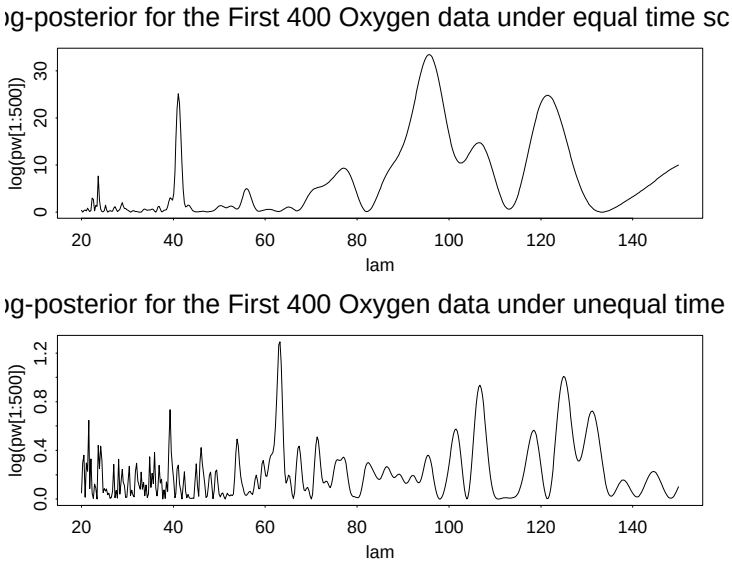

Figure 9.3: *Log-posterior for the first 400 Oxygen data under equally spaced time scale compared with that under unequally spaced time scale*

We first use $p(\omega|Y)$ to study the impact of uncertain timing for the oxygen data used by West (1995). For conventional purposes, we study the posterior of the wavelength instead of the frequency, $p(\lambda|Y)$. We illustrate the effect of different timing scale on $p(\lambda|Y)$ by using both equally and unequally spaced time scale. The equally spaced time scale is based on an interpolation which takes out the data between each adjacent equally spaced time scale and use the average of the adjacent time to impute an equally spaced time scale. Figure 9.3 plots the two log-posteriors. As can be seen, the peak at $\lambda = 96.3$ for the log-posterior under the equal time scale used by most

geologists disappears in the log-posterior under unequally spaced time scale.

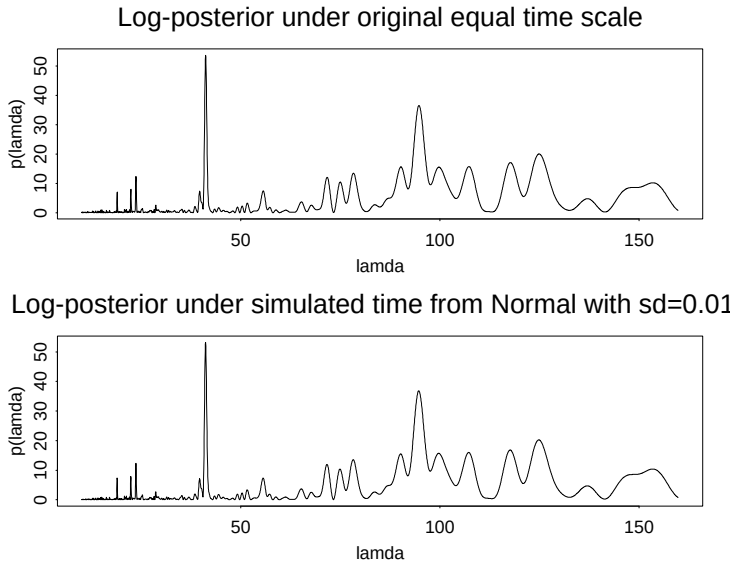

Figure 9.4: *Log-posterior under equally spaced time scale compared with that from a uncertain time scale with error from a Normal distribution with a s.d.=0.017*

Next, we examine the effect of uncertain timing by simulating time from existing time scale. We first simulate time s_i, $i = 1, 2...$ from a normal distribution centered at the original equal time scale $t = 1, 2, ...$ and we require s_i be strictly increasing.

In geological studies, a normal error may represent uncertainties from the calibration process, quality of the core and other sources, etc. This can be represented by an independent normal prior for each time s_i, $p(s_i|t_i, Y) \propto N(t_i, v)$. This normal distribution is truncated at s_{i-1} and s_{i+1} to ensure strict increasing of time. A large variance v means large sampling error for the time. In the oxygen data, each time unit represents 3k years. Therefore, an error of ± 100 years is equivalent to a standard deviation of 0.034 on the unit time scale.

Figure 9.4 shows $p(\lambda|Y)$ for the oxygen data under the equally spaced time scale $t = 1, 2, ...n$ along with $p(\lambda|Y)$ under a time scale simulated from $N(t, 0.034^2)$. The difference between the two log-posteriors is very small. We increase the error range to ± 1000 years, this corresponds to a standard deviation of 0.33 (truncated at s_{i-1}

and s_{i+1}). This means big sampling error of the time. Figure 9.5 shows the two log-posteriors are still similar.

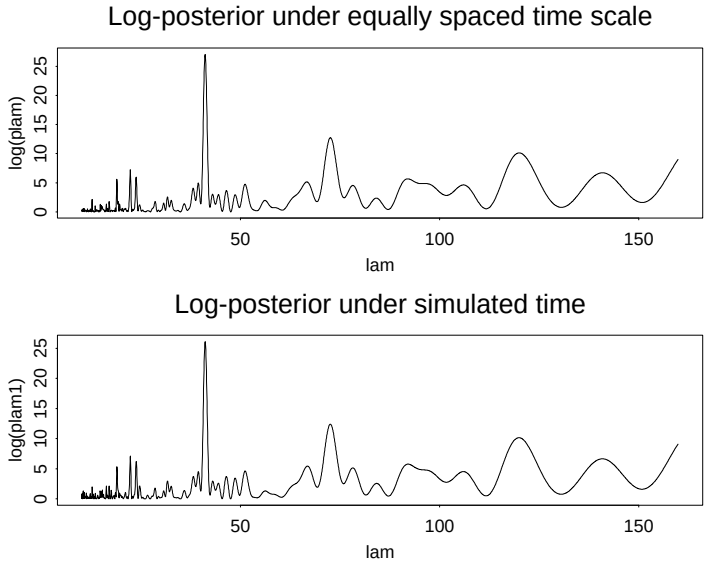

Figure 9.5: *Log-posterior under original equally spaced time scale compared with that from simulated time from* $N(0, 0.33)$

We consider a normal distribution a good representation for the measurement error in the calibration process and the error resulted from low quality of the core. There are other forms of errors to the final timing including the error from uneven sampling of the core which leads to uncertainty in the inter-core depth and errors from the truncation. These types of error can often be represented by a uniform distribution over a specified range, namely $p(s_i) \propto U(s_i - \delta, s_i + \delta)$. This is still an independent prior for each time t_i.

We now experiment the effect on the posterior distribution of the frequency under different time scales simulated from $p(s_i) \propto U(s_i - \delta, s_i + \delta)$ for $i = 1, 2...n$. We use $\delta = 0.3$ which corresponds to ± 900 years of timing error. Figure 9.6 shows no significant changes in the log-posterior from the simulated data.

These interesting results indicate that independent sampling error of time does not have significant impact on the spectral estimate even when the error is considerably

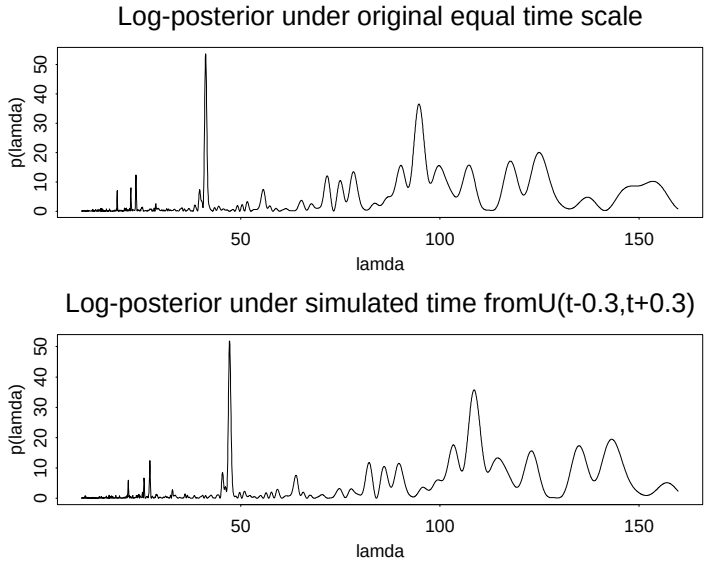

Figure 9.6: *Log-posterior under original time scale and simulated time from* $U(t_i - 0.3, t_i + 0.3)$

large. This is supported by theoretical studies of Moore and Thomson (1991).

Moore and Thomson (1991) considered the case of jittered sampling where samples have been taken at times perturbed from equally spaced grid. Their focus of study was on the impact of jittered sampling on traditional spectral analysis. They showed that when the sampled times are from

$$t_n = n\triangle + \epsilon_n \qquad (n = 0, \pm 1, \pm 2...)$$

where ϵ_n is from a zero mean stationary process, the effect of jittering of the sampling times redistributes the spectral mass of the underlying process, but neither creates nor destroys it. In particular, they derived relationship between the new spectral density function under jittered sampling and the original spectral density function when ϵ_n follows a Gaussian process. They showed that when ϵ_n follows an independent Gaussian process, the relationship between the new spectral density function $f_y(\omega)$ under jittered sampling and the original spectral density $f(\omega)$ is:

$$f_y(\omega) \quad = \quad e^{-\sigma^2\omega^2} f(\omega) + c_0 \tag{9.2}$$

where the constant c_0 is given by

$$c_0 \quad = \quad \int_{-\pi/\triangle}^{\pi/\triangle} (1 - e^{-\sigma^2\omega^2}) f(\omega) d\omega$$

Equation (9.2) indicates the impact of uncorrelated Gaussian jitter is to multiply the spectrum by a roll-off function $e^{-\sigma^2\omega^2}$ and to add the lost weight uniformly across the total frequency range. Significant damage can result if there is weight available for redistribution in $f(\omega)$ at the high frequencies where ω is large relative to $1/^2$. But in general, peak shifting is minimal in the uncorrelated jitter case. Our simulations above support this.

Having concluded that independent sampling error in time does not affect the estimation of frequency much, we study the impact of correlated timing errors on the estimation of spectra. We simulate time s_i from a uniform distribution $U(s_{i-1}, s_{i+1})$ for $i = 1, 2, ...$ subject to $s_{i-1} < s_i < s_{i+1}$. This prior only requires s_i be strictly increasing and it introduces a highly correlated Markov structure to the timing. Figure 9.7 shows that both peaks under the equally spaced time scale shift to the right under dependent simulation error.

The intuition of the above results is clear. When the original time is being "squeezed", the frequency estimate will be higher and the peaks of the posterior distribution for the wavelength will shift to the left. When the original time is being "stretched out", the frequency estimate will be lower and the peaks of the posterior distribution for the wavelength will shift to the right.

Moore and Thomson (1991) also addressed the case of correlated sampling errors. Although exact relationship between new spectral density function under jittered

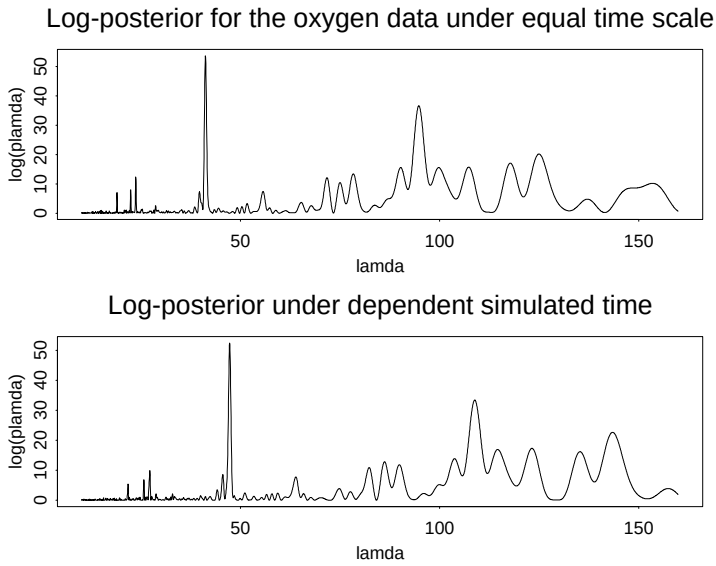

Figure 9.7: *Log-posterior under original time scale and simulated time from* $U(s_{i-1}, s_{i+1})$

sampling $f_y(\omega)$ and the original spectral density $f(\omega)$ can not be developed as in the uncorrelated case, they illustrated potential damages of highly correlated sampling error using simulation studies.

9.3 An Harmonic Model with Uncertain Time

In the previous sections, we demonstrated the impact of timing error on frequency estimation. We have shown that although the effect of uncorrelated timing error is not significant, correlated timing error can result in obvious shifts of peaks in the posterior distribution for the frequency. In this section, we develop estimation procedures for the frequency under uncertain timing for the harmonic model.

A single-frequency harmonic model which incorporates timing uncertainty is

$$y_i \quad = \quad acos(\omega s_i) + bsin(\omega s_i) + \epsilon_i \tag{9.3}$$

where $\epsilon_i \sim N(0, v)$.

Equation (9.3) models the observation y_i as a single harmonic model. Now s_i, $i = 1, ..n$ is a sequence of random variables. Such a model directly counts for timing uncertainty in the model. We conduct inference on frequency and the timing sequence s_i using the observed y_i.

We develop a Bayesian approach to conduct inference on model (9.3). We augment the time sequence $S = \{s_i\}$ to the parameter space $\theta = \{a, b, v, \omega\}$. We can now sample the joint posterior $p(\theta, S|Y)$ by iterating through

- $p(S|\theta, Y)$
- $p(\theta|S, Y)$

We have developed procedures to sample $p(\theta|S, Y)$ in the previous chapters. To sample $p(S|\theta, Y)$, we write

$$p(S|\theta, Y) \propto p(S)p(Y|S, \theta)$$

Here we assume that S is independent of the model parameters θ. Because of the nonlinear structure of $p(Y|S, \theta)$, direct sampling of the joint posterior $p(S|\theta, Y)$ is not feasible. Therefore, we further break the joint posterior $p(s_1, s_2, ...s_n|\theta, Y)$ into $p(s_i|S^i, \theta, Y)$ for $i = 1, 2..n$ where $S^i = \{s_1, s_2, ..s_{i-1}, s_{i+1}, ..s_n\}$. Since $p(Y|S, \theta)$ can be factored into products of individual terms $p(y_i|s_i, \theta)$, the conditional distribution $p(s_i|S^i, \theta, Y)$ can be written as

$$p(s_i|S^i, \theta, Y) \propto p(s_i|S^i)p(y_i|s_i, \theta)$$

This can be viewed as the product of the prior for each s_i given other s_j and the likelihood for y_i given S_i and θ. We now study the choice of prior $p(s_i|S^i)$.

As discussed earlier, one choice of the prior $p(s_i|S^i)$ is the independent uniform distribution $p(s_i|S^i) = U(t_i - \delta, t_i + \delta)$ where t_i is the equally spaced time. This may represent truncation errors from measurement. A more flexible prior is the

dependent uniform distribution $p(s_i|S^i) = U(s_{i-1}, s_{i+1})$. This prior only requires a strict increasing order of the time. Some more sophisticated priors can be constructed based on knowledge of the collecting process of the core and the calibration process. For example, in the plot of oxygen data series, it can be seen that the top of each quasi-cycle is not smooth, but rather flat. A cosine wave of the equally spaced time will not capture such behaviour. However, it can be approximated by a cosine wave under a new time scale. Such new time scale within each period looks like the top plot in Figure 9.8. It can be seen that after such time transformation, a smooth cosine wave will look like the "flat-top" wave in the bottom of Figure 9.8.

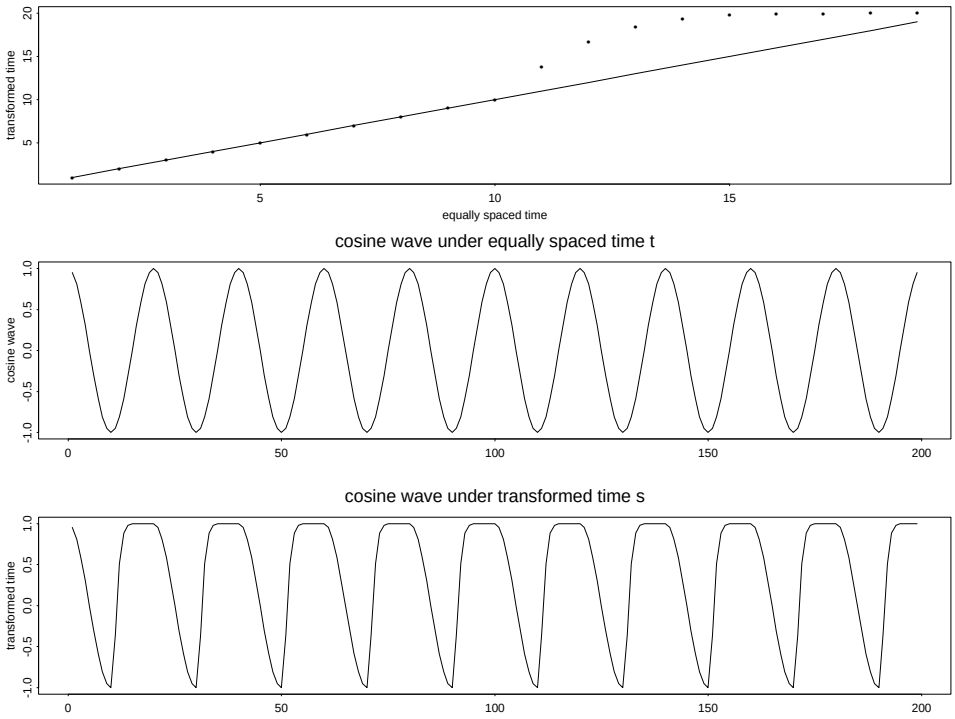

Figure 9.8: *An example of true time vs. equally spaced time*

We discuss a possible prior for such timing scales. Let $T = \{t_1, t_2, ...t_n\}$ denote the equally spaced time. We assume the observation time is subject to normal errors around the equally spaced time T. So we would like to have a normal prior on the

observation times S. A few things we know about S are: $s_1, s_2, ..s_n$ has to be positively increasing; s is distributed with a mean function of t; the function s(t) is a periodic function. so $s(t+\lambda) = s(t)$. To have normal priors, we need to have unbounded variables which can be both positive and negative. A possible transformation of S to satisfy this requirement is to assign the log of $s(t)/(\lambda - s(t))$ to have a normal distribution; notice that $0 < s(t) < \lambda$ since $s(t+\lambda) = s(t)$. The mean of the normal distribution is

$$m(t) = log\frac{m^*(t)}{\lambda - m^*(t)}$$

where $m^*(t) = t \bmod \lambda$.

In summary, let $x(t) = log\frac{s(t)}{\lambda - s(t)}$. The prior for $p(s_i|S^i)$ can be specified by

$$x(t) \sim N(m(t), v)$$

for some variance v. If v is small, $x(t)$ is close to $m(t)$ and so $s(t)$ is near t mod λ and the deformation is small. If v is large, $x(t)$ is far from $m(t)$ and so $s(t)$ is far from t mod λ and the deformation is large.

The last constraint is that s(t) should be increasing in t, so $x(t_{i-1}) < x(t_i) < x(t_{i+1})$. This leads to a truncated normal for $p(x(t_i)|other x(t_i), \theta)$. so

$$\begin{aligned} x(t) &\sim N(m(t), \omega) && \mathrm{x(t_{i-1}) < x(t_i) < x(t_{i+1})} \\ &\sim 0 && \text{otherwise} \end{aligned}$$

We can easily sample from a truncated normal distribution, and since $p(y_i|s_i, \theta)$ is easy to evaluate, we can use a Metropolis algorithm to sample the posterior

$$p(s_i|S^i, \theta, Y) \propto p(s_i|S^i)p(y_i|s_i, \theta)$$

We can use $p(s_i|S^i)$ as a candidate generating distribution to deliver a candidate sample s_i^*. The sample is either accepted or rejected according to the Metropolis acceptance ratio:

$$\alpha(s_i^*, s_i) = min\{1, \frac{p(y_i|s_i^*, \theta)}{p(y_i|s_i, \theta)}\}$$

where s_i is the last sample from $p(s_i|S^i, \theta, Y)$. This ratio depends only on the evaluation of the conditional normal densities determining the involved likelihood ratio which can be easily computed.

In summary, the simulation approach to formally address the problem of uncertain timing can be developed by a Gibbs Sampler iterating through

- $p(S|\theta, Y)$
- $p(\theta|S, Y)$

The sampling of $p(S|\theta, Y)$ can be further broken down to sample $p(s_i|S^i, \theta, Y)$ $i = 1, 2, ..., n$, iteratively. Each $p(s_i|S^i, \theta, Y)$ can be sampled by a Metropolis algorithm using the prior distribution as the candidate generating distribution.

9.4 Application

We now apply the harmonic model with unknown timing to analyze the Oxygen data and the EEG data studied before.

We first apply the harmonic model with unknown timing to the oxygen data. As discussed earlier, the oxygen data is subject to severe timing error. Possible causes of such error include errors in the sampling of the core including the quality of the core, and the error from the unequal sampling of the core and gaps between cores

which resulted in the uncertainty of the inter-core depth. Other errors come from carbon-14 calibration of the core, and the error from the interpolation of unequal time scale to equal time scale. Possible priors discussed earlier include the independent error and the dependent error. We now compare the frequency estimate using a four-frequency harmonic model without timing error, with independent timing error and with dependent timing error. The focus is on how the introduction and different choices of priors of timing error affects the estimation of frequency and what inference on timing can be drawn from the data using different priors.

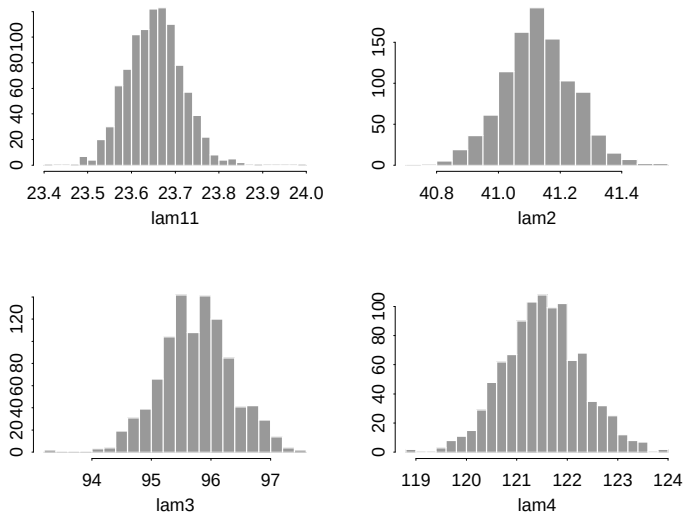

Figure 9.9: *Histograms of the opsteriors for wavelengths using a four-frequency model without timing error*

We first apply a four-frequency harmonic model without timing error to the first 400 observations of the oxygen data. This is basically a repeat of the analysis in section 8.3.2 of Chapter 8. The Gibbs Sampler was run for 2000 iterations. Figure 9.9 shows the histograms of the four wavelengths. The rejection rates for each Metropolis chain are: 41%, 39%, 28%, and 31%. The means and variances of the four wavelengths are $\mu_1 = 23.62$, $v_1 = 0.058$, $\mu_2 = 41.13$, $v_2 = 0.012$, $\mu_3 = 95.75$, $v_3 = 0.376$, $\mu_4 = 121.52$, $v_4 = 0.59$. These four wavelengths corresponds to the precession (19-23k

years), obliquity of the earth's orbit (40-42k years) and periodicities in eccentricity (95-100k years and 120-130k years).

Second, we apply a four-frequency harmonic model with independent timing error to the same oxygen data. Namely, the prior for each s_i is $p(s_i) = U(t_i - 0.3, t_i + 0.3)$ for $t_i = 1, 2....n$. Figure 9.10 shows the trajactory plot of the model parameters from the Gibbs Sampler. Figure 9.11 shows the histograms of the four wavelengths. The means and variances of the four wavelengths are $\mu_1 = 23.65$, $v_1 = 0.0048$, $\mu_2 = 41.13$, $v_2 = 0.02$, $\mu_3 = 95.79$, $v_3 = 0.387$, $\mu_4 = 121.52$, $v_4 = 0.62$. These statistics are very close to those without timing error. It can be seen that the introduction of independent sampling error to time does not change the estimates of the frequencies much. A plot of the inferred time versus the equally spaced time scale in Figure 9.12 shows small variations of the inferred time around the equally spaced time. The scale of the variation is within ± 0.04, which corresponds to ± 120 years of variation. Comparing with the range used in the prior ± 0.3, this means the data is suggesting a smaller amount of variation in the sampling error of time given the model.

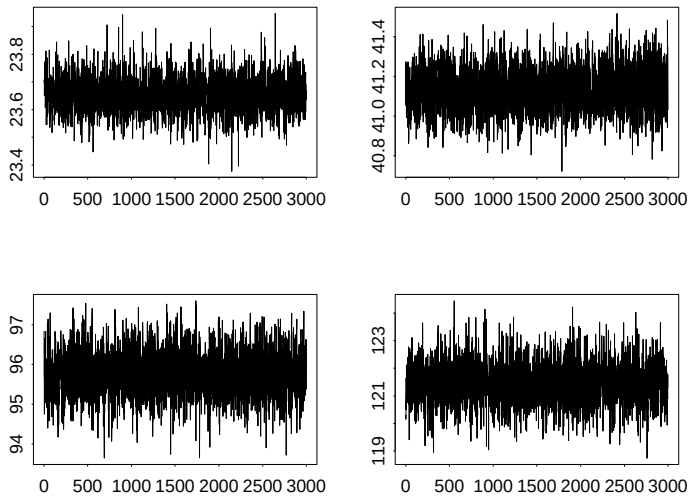

Figure 9.10: *Trajectory plots of simulations of the wavelengths under independent timing error from* $U(s_i - 0.3, s_i + 0.3)$

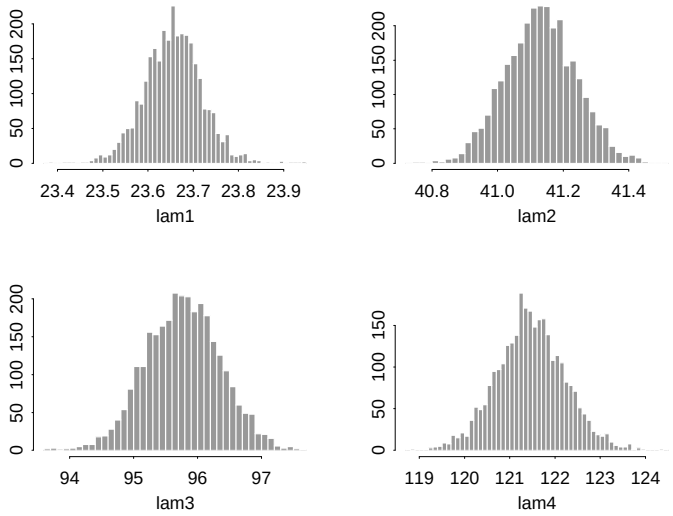

Figure 9.11: *Posterior distributions of the wavelengths using a four-frequency model with independent timing error*

Third, we apply a four-frequency harmonic model with dependent timing error to the same oxygen data. Namely, the prior for each s_i is $p(s_i) = U(s_{i-1}, s_{i+1})$ for $i = 1, 2....n$. As stated before, this general prior only requires the timing to be strictly increasing. This introduces a highly correlated Markovian structure for the prior on the time.

60000 iterations was run from the Gibbs Sampler with everything 30 iterations saved. This forms a chain of 2000 samples to be used for inference. Figure 9.13 shows the trajectory plot of the simulations. The acf plotted for each λ indicates the adjacent samples are highly correlated until lag 30 for most wavelengths. As these samples are every 30 iterations from the original Gibbs Sampler, this indicates large autocorrelations exist for lag 900.

This slow convergence problem is similar to what's observed in the Stochastic Volatility Model in the first part of this thesis. Yet the Multi-move Gibbs Sampler which considerably speeds up convergence in the Stochastic Volatility Model can not be used here because of the complex non-linear structure of the wavelength.

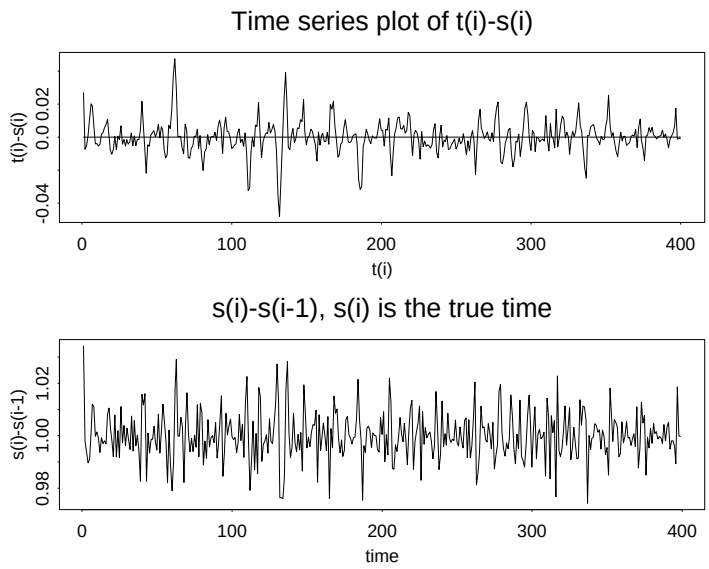

Figure 9.12: *Top: Time series plot of true time minus equally spaced time. Bottom: Time series plot of the difference of adjacent true time.*

Innovative simulation procedures need to be developed to improve convergence for the highly correlated Markovian structure of the time in this model before reliable inference of the model parameters can be made.

The histogram of the wavelengths are plotted in Figure 9.15. The means and variances of the four wavelengths are $\mu_1 = 20.11$, $v_1 = 0.005$, $\mu_2 = 41.27$, $v_2 = 0.024$, $\mu_3 = 94.75$, $v_3 = 0.48$, $\mu_4 = 120.88$, $v_4 = 1.56$. It's clear that the high frequency λ_1 has shifted significantly from near $\mu_1 = 23.62$ without timing error to near $\mu_1 = 20.11$. The variance for μ_1, however, decreases from $v_1 = 0.058$ to $v_1 = 0.005$. Other wavelengths also shift slightly and the variance for λ_4 increases considerably. Therefore, the introduction of highly dependent sampling error to time changes the estimates of the high frequency. A plot of the inferred time versus the equally spaced time scale in Figure 9.16 shows obviously jittering around the equally spaced time scale. The plot of the difference of the adjacent inferred time $s_i - s_{i-1}$ has shocks at various times which may be compared with any external knowledge on the sampling error of the core, etc. The mean of the sequence of the difference has mean 0.989.

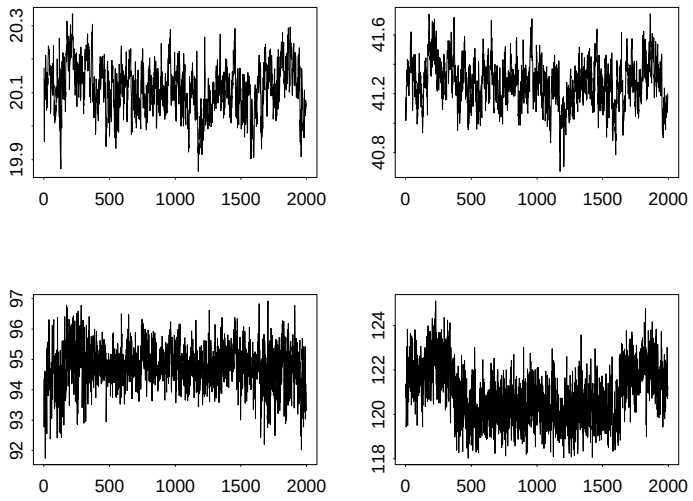

Figure 9.13: *Trajectory plots of simulations of the wavelengths under dependent timing error*

Table 9.1: Means and variances of the four wavelengths under no timing error, independent timing error and dependent timing error

	λ_1	λ_2	λ_3	λ_4
No timing error	23.62 (0.058)	41.13 (0.012)	95.75 (0.376)	121.52 (0.59)
Independent timing error	23.65 (0.048)	41.13 (0.02)	95.79 (0.387)	121.52 (0.62)
Dependent timing error	20.11 (0.05)	41.27 (0.024)	94.75 (0.48)	120.88 (1.56)

Table 9.1 lists the mean and variance of the four wavelengths under no timing error, independent timing error and dependent timing error. It can be seen that the mean of the wavelengths are very similar under no timing error and independent timing error. Yet the means shift considerably for dependent timing error especially for λ_1.

The oxygen data was the motivating example to study the impact of timing uncertainty on the estimation of frequencies. Alternatively, we study the impact of uncertain timing from a different point of view using the EEG data.

Timing uncertainty is not a major problem in the collection of the EEG data. But the study of West (1996) suggested the existence of time-varying frequency in some

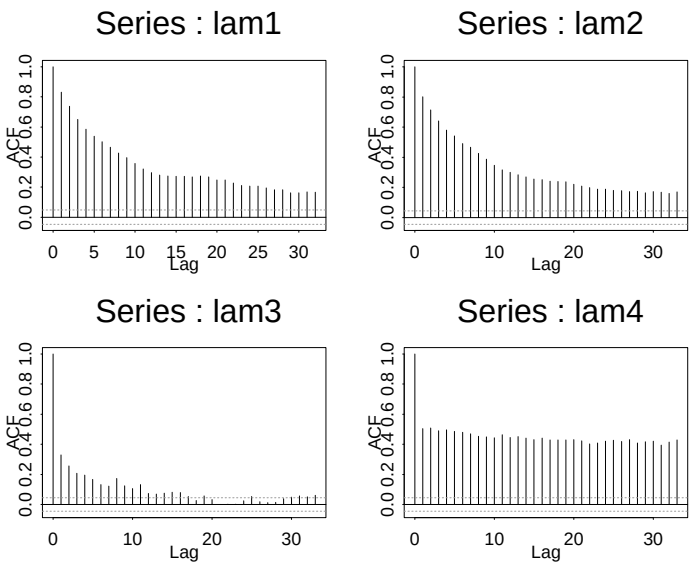

Figure 9.14: *ACF of the samples from the oxygen data with dependent prior*

EEG data series. West has developed time-varying frequency models in the time domain to analyze the data. Here we apply a fixed-frequency harmonic model and conduct inference on the time given the data. So instead of studying how frequencies change through time under fixed time scale, we study how time evolves under fixed frequencies.

Figure 9.17 shows the posterior distribution of the wavelength for all 300, the first 150 and the last 150 EEG data. It's obvious that there is one important wavelength changing through time from $\lambda = 5.1$ to $\lambda = 6.2$.

We first apply a one-frequency harmonic model with fixed time to the data. Figure 9.18 shows the histogram of the wavelength, the amplitudes and the variance. The histogram of the frequency has two clear peaks which correspond to the two peaks in the posterior distribution of the wavelength for the 300 data sets.

We then assign the same depend prior used for the oxygen data to the time. Namely, $p(s_i) = U(s_{i-1}, s_{i+1})$ for $i = 1, 2....n$. Unlike the Oxygen data, there is less uncertainty in the EEG data y_i. This causes the Gibbs Sampler to converge

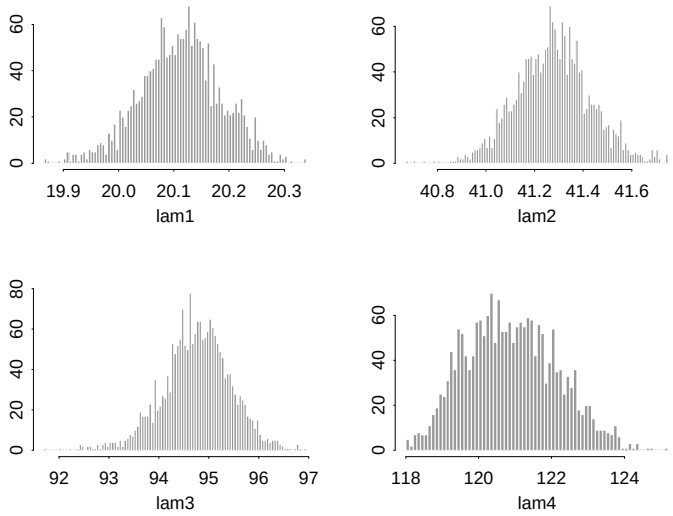

Figure 9.15: *Histogram of the wavelengths for the Oxygen data under dependent timing error*

even slower. Figure 9.19 shows the simulation trajectory plots of the wavelength and other model parameters. They represent simulation chains of $600,000$ iterations after taking out every 20 iterations. From the trajectory plot of the wavelength, it can be seen the chain is converging very slowly. Figure 9.20 shows the simulation trajectory plots of the times at $t = 50$,$t = 100$, $t = 150$, $t = 200$, $t = 250$ and $t = 300$, they also indicate slow convergence.

Reliable inference on the parameters can not be achieved with the existence of slow convergence. Yet it is still interesting to compare the posterior distribution of the wavelength under no timing uncertainty and with timing uncertainty. Figure 9.21 shows the histogram of the wavelength and other model parameters. The 95% for the wavelength under uncertain time is $(4.687, 5.282)$ whereas the two wavelengths without timing error are centered at 5.02 and 5.27.

Finally, we study the inference on the time for the EEG data. As stated earlier, timing uncertainty is not a major problem in the collection of the EEG data. Rather, this EEG data series exhibits time-varying frequency. Using a fixed-frequency har-

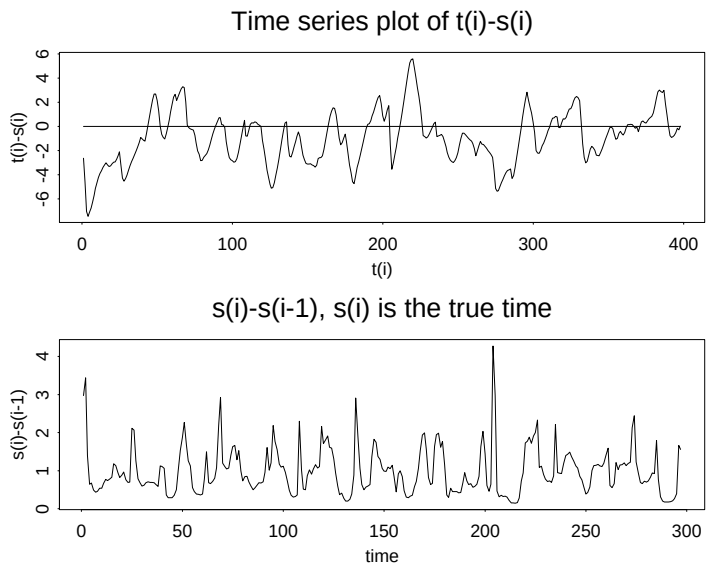

Figure 9.16: *Top: Time series plot of true time minus equally spaced time. Bottom: Time series plot of the difference of adjacent true time.*

monic model and uncertain time, we study how time evolves under fixed frequencies. Since the wavelength increases through time, we expect to see an increasing trend on the difference of adjacent inferred time $t_i - s_i$ in the second part of the data under the fixed-wavelength model. This is confirmed by the second part of the plot of the difference of adjacent inferred time $t_i - s_i$ in Figure 9.23.

9.5 Conclusion

In this chapter, we discussed the issue of timing uncertainty in time series analysis. In particular, we studied the impact of uncertain timing on the estimation of frequencies in harmonic analysis. Our empirical study on the oxygen data indicates that indecently timing error has little effect to the estimation of the frequencies whereas highly dependent timing error has large effect to the estimation of frequencies. These findings are consistent with theoretical discussions on the impact of sampling error to the estimation of frequencies in the spectral analysis literature. In addition, we develop harmonic models which incorporate timing uncertainty. We

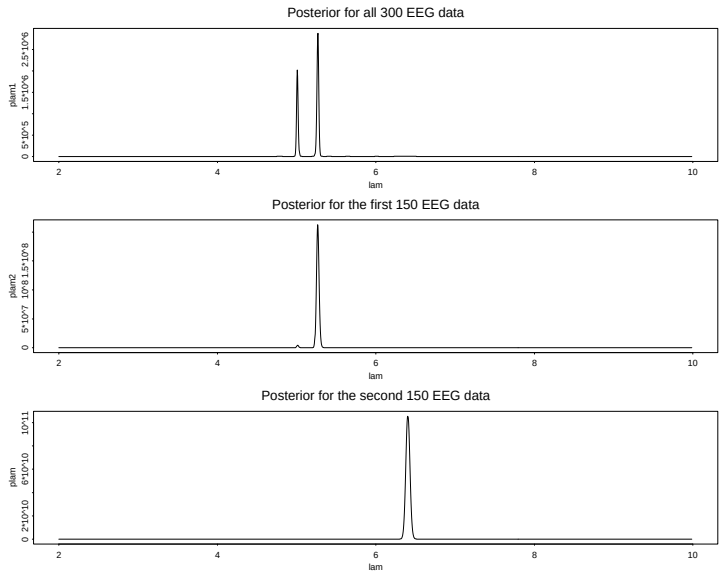

Figure 9.17: *Posterior distribution for the wavelength for all 300, the first 150 and the last 150 EEG data*

develop simulation-based inference methods for the estimation of frequencies under uncertain time. These methods were applied to the study of the oxygen data and EEG data. We illustrated the estimation of frequencies under different prior distributions for the timing uncertainty. Further development of prior distributions for the time is of major interest in future work. Also, innovative simulation procedures need to be developed for highly dependent prior distribution to faster convergence.

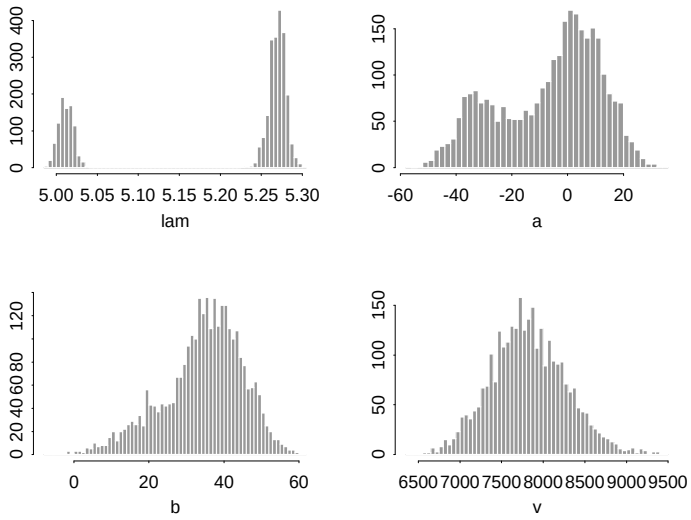

Figure 9.18: *Histogram of the model parameters for the EEG data under equally spaced time*

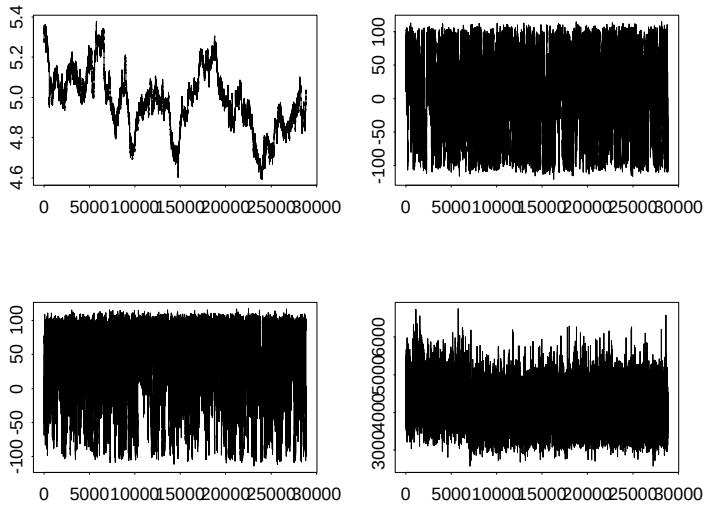

Figure 9.19: *Simulation trajectory plots for λ, a,b,v*

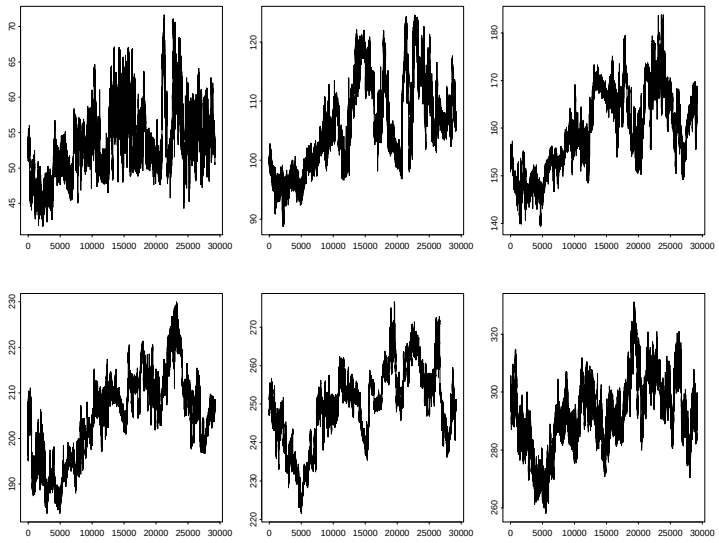

Figure 9.20: *Simulation trajectory plots for time* $t = 50$,$t = 100$, $t = 150$, $t = 200$, $t = 250$ *and* $t = 300$

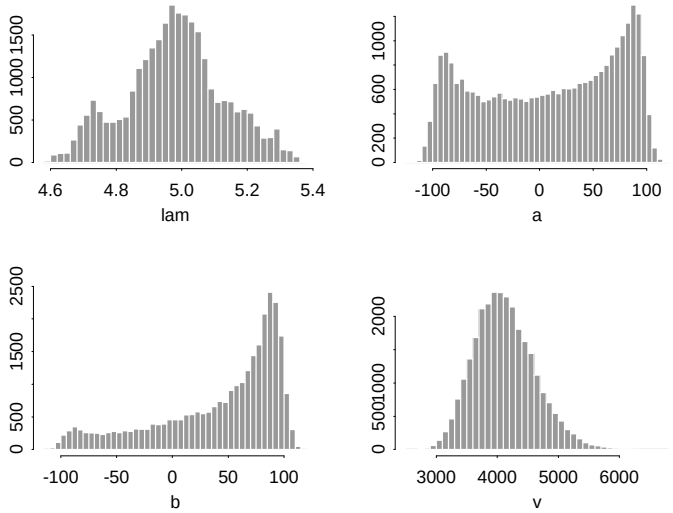

Figure 9.21: *Histogram of wavelength and other model parameters*

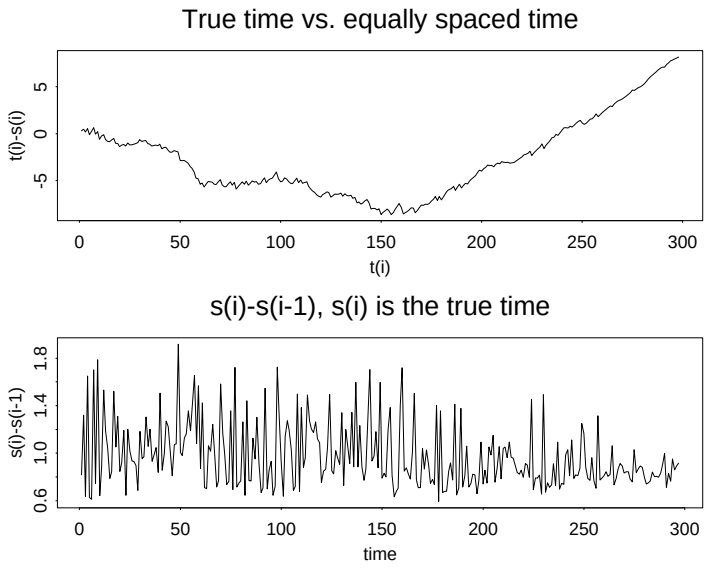

Figure 9.22: *Top: Time series plot of true time minus equally spaced time. Bottom: Time series plot of the difference of adjacent true time.*

Chapter 10

Conclusions and Future directions

In this part of the disseration, we develop Bayesian analysis on spectral estimation in the frequency domain of time series analysis. This disseration provides an efficient and unified simulation procedure for frequency estimation under a harmonic model framework. The focus of the frequency estimation in a harmonic model is the sampling of the posterior distribution of the frequencies. We developed a mode-based Metropolis sampling approach to tackle the challenging problem of sampling highly multi-modal and high dimensional posterior distribution of the frequency. The method can be generalized to sample other multi-modal posterior distributions. The Gibbs Sampler approach developed in this disseration for frequency estimation allows the development of more sophosticated dynamics for frequency domain time series analysis.

Many interesting directions for further developments of the harmonic models developed in this disseration exist. An immediate extension is models that allow time-varying amplitudes and phases. Many cyclical time series data exbihit time-varying amplitudes and phases. For example, the EEG data studied extensively in this disseration exibihits time-varying amplitudes in some channels. We have studied such features in an alternative way using fixed-frequency model with uncertain time. Dynamic models in the time domain that allow time-varying amplitudes and phases have been developed by Prado and West (1997), for example. Developments of such models in the frequency domain is still an untouched area of research.

Motivated by timing issues in geological time series study, we developed a harmonic model which incorporates timing uncertainty. This model is a good illustration

of the use of Gibbs Sampler for frequency estimation in complex models. The model directly accounts for timing uncertainty in the estimation of frequencies. This allows the study of the impact on timing uncertainty on the estimation of frequencies. At the mean time, inference on time given observed data can also be achieved. Geology has come a long way without an accurate time scale. Knowledge on possible causes of timing uncetainty in the sampling and measurement procedures usually exist, the model developed in this disseration incorporates such knowledge into the prior distribution and conducts formal inference on timing and frequency estimation using such knowledge. We experimented with different choices of priors which represent different types of sampling errors in time. Our results indicate choice of priors have obvious impact on the inference on time. When prior knowledge indicates independent sampling errors of time, our study shows such timing error does not have significant impact of the estimation of frequencies especially when the frequencies are not very high. If prior knowledge indicates dependent sampling errors of time, our study shows such timing error may change the estimation of frequencies dramatically especially when the timing errors are highly correlated. Also, our study found very slow convergence of the Gibbs sampler in models with highly correlated timing errors. This is similar to the slow convergence problem in the stochastic volatility models when volatilities are highly correlated. In that case, we were able to approximate the nonlinear model by mixtures of dynamic linear models and apply existing simulation methods for dynamic linear models to speed up convergence. However, the same technique can not be transferred to the nonlinear structure presented in the harmonic model. Innovative simulation techniques need to be developed.

The model with timing uncertainty is a good lead to a more general topic, time deformation contexts which model the usual dynamic processes from a different point of view, the dynamics of the evolution of time. This has both theoretical impact on

the link between linear stationary process and nonlinear nonstationary process and practical applications where the inference on timing is of interest. For example, the operational time which relates to news releases in the financial market versus the calendar time when prices of various financial instruments are recorded. In the study of the application of time deformation models, identifying the actual deformation process of time is crucial. The simulation methods developed in our model lay a basis for future research.

Bibliography

[1] Amin, K.L. and Ng, V. (1993), "Equilibruim Option Valuation with Systemiatic Stochastic Volatility", *Journal of Finance* 48, 881-910.

[2] Andersen, T.G. (1993), "Stochastic Autoregressive Volatility: A Framework for Volatility Modeling", Working Paper 161, Northwestern University, Dept. of Finance, J.L.Kellogg Graduate School of Management. (To appear in Mathermatical Finance.)

[3] Andersen, T., and Sorensen, B. (1993), "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study", Working paper, Northwestern University, J.L. Kellogg Graduate School of Management.

[4] Andersen, T.G. (1994), Stochastic Autoregressive Volatility: A Framework for Volatility Modeling", *Mathermatical Finance* 4, 75-102.

[5] Andersen, T.G. and Sorensen, B. (1994), "Estimation of a Stochastic Volatility Model: a Monte Carlo Study", Working paper, Northwestern University, J.L. Kellogg Graduate School of Management.

[6] Andersen, T.G. and Sorenson, B.E. (1994), "GMM and QML Asymptotic Standard Deviations in Stochastic Volatiity Models: A Response to Ruiz. Working Paper 189, Dept. of Finance, Kellogg School, Northwestern University.

[7] Anderson, T.G. (1995), "Return Volatility and Trading Volume: an Information Flow Interpretation of Stochastic Volatility, *Journal of Finance,* Forthcoming.

[8] Ball, C (1993), "A Review of Stochastic Volatility Models with Application to Option Pricing." Proceedings of Conference on Financial Innovation: 20 Years of Black/Scholes and Merton (November 1993), 608-644.

[9] Bates, D.S. (1995), "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in PHLX Deutschemark Options", *Review of Financial Studies,* Forthcoming.

[10] Bollerslev, T. Chou, R. and Kroner, K. (1992), "ARCH modeling in Finance", *Journal of Econometrics* 52, 5-59.

[11] Bollerslev, T. (1993), "Long Memory in Stochastic Volatility", Discussion paper, Northwestern University.

[12] Breidt, F.J. (1996), "A Threshold Autoregressive Stochastic Volatility Model," Working paper, Iowa State University.

[13] Breidt, F.J., Crato, N., and de Lima, P. (1994), "On the Detection and Estimation of Long memory in Stochstic Volatility", To appear in *Journal of Econometrics.*

[14] Breidt, F.J., de Lima, P., and Crato, N. (1994), "Modeling Long Memory Stochastic Volatility", Working Papers in Economics No. 323, Johns Hopkins University, Iowa State University.

[15] Bretthorst, G.L. (1988), "Bayesian Spectrum Analysis and Parameter Estimation", Lecture notes in statistics. Springer-Verlag, New York.

[16] Brillinger, D.R. (1972), "The Spectral Analysis of Stationary Interval Functions," *Proceedings of the 6th Berkeley Symposium on mathematical statistics and probability,* 483-513.

[17] Brillinger, D.R. (1983), "Statistical Inference for Irregularly Observed Data", Lecture notes in statistis, 25, 38-57. Springer-Verlag, New York.

[18] Canina, L. and Figlewski, S. (1993), "The Informational Content of Implied Volatility", *Review of Financial Studies* 6, 659-682.

[19] Carter, C.K. and Kohn, R. (1994)," On Gibbs Sampling for State Space Models", *Biometrika* 81, 541-553.

[20] Casella, G and George, E. (1992)," Explaining the Gibbs Sampler", *American Statistical Association* 46, No. 3. 167-174.

[21] Chesney, M. and Scott, L.O. (1989), "Pricing European Currency Options: A Comparision of the Modified Black-Scholes Model and a Random Variance Model", *Journal of Financial and Qualitative Analysis* 24, 267-284.

[22] Christie, A.A. (1982), "The Stochastic Behavior of Common Stock Variances: Value, Leverage, and Interest Rate Effects", *Journal of Financial Economics* 10, 407-432.

[23] Clark, P.K. (1973), "A Subordinated Stochastic Process Model with Finite Variances for Speculative Prices", *Econometrica,* 41, 135-156.

[24] Clewlow, L. and Xu, X. (1993), "The Dynamics of Stochastic Volatility", Discussion Paper, University of Warwick.

[25] Comte, F. and Renault, E. (1995), "Long Memeory Continuous Time Stochastic Volatility Models", Paper presented at the HFDF-I Conference, Zurich.

[26] Crnkovic and Drachman (1996), "A Universal Tool to Discriminate among Risk Measurement Techniques", To appear in *Risk*.

[27] Danielsson J. (1993), "Multivariate Stochastic Volatility", Working paper, University of Iceland, Dept. of Economics.

[28] Danielsson, J. (1996), "Stochastic Volatility in Asset Prices: Estimation With Simulated Maximum Likelihood", *Journal of Econometrics*, forthcoming.

[29] Day, T. E. and Lewis, C. (1992)," Stock Market Volatilities and the Information Content of Stock Index Options", *Journal of Econometrics* 52, 267-287.

[30] Dupire, B. (1992), "Arbitrage Pricing with Stochastic Volatility." Working Paper, Proceedigns of AFFI Conference, Paris.

[31] Eisenberg, L.K. and Jarrow, R.A. (1991), "Option Pricing With Random Volatilities in Complete Markets," Federal Reserve Bank of Atlanta Working Paper 91-16.

[32] Engle, R.F. and Lee, G.G.J. (1994), "Estimating Diffusion Models of Stochastic Volatility", Working Paper: Dept. of Economics UCSD.

[33] Fruhwirth-Schnatter, S. (1994), "Data augmentation and dynamic linear models",*Journal of Time Series Analysis* 15, 183-202.

[34] Gallant, A.R., Hsieh, D., and Tauchen, G. (1995), "Estimation of Stochastic Volatility Models with Diagnostics", Working paper, Duke University.

[35] Gardner, W.A. (1987), "Statistical Spectral Analysis", Prentice Hall, New York.

[36] Gelman, A. and D.B. Rubin (1992), "Inference from Iterative Simulation using Multiple Sequences", *Statistical Science* 7, 4, 457-511.

[37] Geweke, J. (1994), Comment on "Bayesian Analysis of Stochastic Volatility Models", *Journal of Business and Economic Statistics* 122, 397-399.

[38] Ghysels, E. and Jasiak, J. (1993), "Stochastic Volatiility and Time Deformation: An Application of Trading Volume and Leverage Effects", Discusssion paper, C.R.D.E., Montreal.

[39] Ghysels, E., Harvey, A., and Renault, E. (1995), "Stochastic Volatility", Working paper, prepared for *Handbook of Statistics* Vol. 14.

[40] Ghysels, E., Gourieroux, C., and Jasiak, J. (1995), "Trading Patterns,Time Deformation and Stochastic Volatility in Foreign Exchange Markets", Paper presented at the HFDF Conference, Zurich.

[41] Gilks, W.R. Richardson, S. and Spieglhalter, D.J. (1996), "Markov Chain Monte Carlo in Practice", Chapman and Hall, London.

[42] Harvey, A.C. (1993), "Long Memory in Stochastic Volatility", Discussion Paper 10, London School of Economics, Dept. of Statistics.

[43] Harvey, A.C., Ruiz, E., and Shephard, N. (1994), "Multivariate Stochastic Variance Models", *Review of Economic Studies* 61, 247-264.

[44] Harvey, A.C., and Shephard, N. (1993), "The Econometrics of Stochastic Volatility", Working paper, London School of Economics, Dept. of Statistics.

[45] Harvey, A.C., and Shephard, N. (1996), "Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns", *Journal of Business and Economic Statistics* (forthcoming).

[46] He, H. (1993), "Option Prices with Stochastic Volatilities: An Equilibrium Analysis", Discussion Paper, University of California, Berkeley.

[47] Heston, S.L. (1993), "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options", *Review of Financial Studies* 6, 327-343.

[48] Hull, J. and White, A. (1987), "The Pricing of Options on Assets With Stochastic Volatilities", *Journal of Finance* 42, 281-300.

[49] Hull, J. and White, A. (1988), "An Analysis of the Bias in Option Pricing Caused by a Stochastic Volatility." *Advances in Futurues and Options Research* 3, 29, 61.

[50] Jacquier, E., Polson, N.G., and Rossi, P.E. (1993), "Stochastic Volatility: Univariate and Multivariate Extensions", Under revision at *the Journal of Econometrics.*

[51] Jacquier, E., Polson, N.G., and Rossi, P.E. (1995), "Priors and Models for Multivariate Stochastic Volatility", Unpublished manuscript, University of Chicago,

Graduate School of Business.

[52] Jacquier, E., Polson, N.G., and Rossi, P.E. (1994), "Bayesian Analysis of Stochastic Volatility Models (With discussion)", *Journal of Business and Economic Statistics* 12, 371-417.

[53] Jacquier, E., Polson, N.G., and Rossi, P.E. (1995), "Priors and Models for Multivariate Stochastic Volatility", Working paper, Johnson School of Management, Cornell University.

[54] Jacquier, E., Polson, N.G., and Rossi, P.E. (1995), "Multivariate and Prior Distributions for Stochastic Volatility Models", Discussion Paper CIRANO.

[55] Jacquier, E., Polson, N.G., and Rossi, P.E. (1995), "Stochastic Volatility: Univariate and Multivariate Extensions", Rodney White Center for Financial Research Working Paper 19-95, The Wharton School, University of Pennsylvania.

[56] Jaynes, E.T. (1983), "Papers on Probability, Statistics and Physics", a reprint collection, D. Reidel, Dordrecht-Holland.

[57] Jaynes, E.T. (1987), "Bayesian Spectrum and Chirp Analysis", *Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems* 1-37.

[58] Jones, C. (1989), "Foreign Exchange Option Pricing", Barra Newsletter.

[59] Jorion, P. (1995), "Predicting Volatility in the Foreign Exchange market", *Journal of Finance* 50, 213-245.

[60] Kay, S.M. and Maple S.L. (1981) "Spectrum Analysis- A Modern Perspective", *Proceedings of the IEEE* 69, 11, 1380-1419.

[61] Kim, S. and Shephard (1994), "Comment on 'Bayesian Analysis of Stochastic Volatility Models' by Jacquier, Polson and Rossi", "Journal of Economics and Business Statistics 12, 371-417.

[62] Kroner, K. (1995), "Creating and Using Volatility Forecasts", Working paper, BZW Global Investors.

[63] Lamoureux, C.G. and Lastrapes, W.D. (1993), "Forecasting Stock-Return Variance: Toward an Understanding of Stochastic Implied Volatilities", *Review of Financial Studies* 6, 293-326.

[64] Liu, J. W.H. Wong and A. Kong (1994), "Covariance Structure of the Gibbs Sampler with Applications to the Comparison of Estimators and Augmentation Schemes", *Biometrika* 81, 27-40.

[65] Mahieu, r. and P. Schotman (1994), "Stochastic Volatility and the Distribution of Exchange Rate News", Working paper: University of Limburg.

[66] Martin, P (1995), "Blunders that Bust the Bank", *Financial Times* March 23, p24.

[67] Meddahi, N. and Renault, E. (1995), "Aggregations and Marginalisations of GARCH and Stochastic Volatility Models", Discussion Paper, GREMAQ.

[68] Melino, A., and Turnbull, S. (1990), "Pricing Foreign Currency Options With Stochastic Volatility", *Jounal of Econometrics* 45, 7-39.

[69] Moore, M.I. and P.J. Thomson (1991), "Impact of Jittered Sampling on Conventional Spectral Estimates", *Journal of Geophysical Research* 96, 18,519-18,526.

[70] Nelson, D. (1991), "Conditional Heteroscedasticity in Asset Pricing: a New Approach", *Econometrica* 59, 347-370.

[71] Park, J. (1991), "Envelope Estimation for Quasi-Periodic Geophysical Signals in Noise: a Multitaper Approach", *Statistics in the Environmental and Earth Sciences,* edited by Walden, A.T. and Guttorp, P. 189-219, Edward Arnold, London.

[72] Park, J. and Maasch, K.A. (1993), "Plio-Pleistocene Time Evolution of the 100-hyr Cycle in Marine Paleoclimate Records", *Journal of Geophysics Research*, 98, 447-461.

[73] Parzen, E.M. (1983), "Time Series Analysis of Irregularly Observed Data", Lecture notes in statistics, 25, Springer-Verlag, New York.

[74] Pasorello, S., Renault, E., and Touzi, N. (1993), "Statistical Inference for Random Variance Option Pricing", Discussion Paper, CREST

[75] Priestley, M.B. (1981), Spectral Analysis and Time Series, Academic Press, New York.

[76] Raftery, A. and Lewis, S. (1992), "How Many Iterations of the Gibbs Sampler?" Bayesian Statistics 4, (eds: J.O. Berger, J.M. Bernardo, A.P. Dawid and A.F.M. Smith), Oxford University Press, London.

[77] Ruddiman, W.F. McIntyre, A. and Raymo, M. (1986), "Paleoenvironment Results from North Atlantic Sites 607 and 609", *Institutional Reports, DSDP*, 84, U.S. Government printing office, Washington, 855-878.

[78] Ruiz, E. (1994), "Quasi-Maximum Likelihood Estimation of Stochastic Variance Models", *Journal of Econometrics* 63, 284-306.

[79] Scott, L.O. (1987), "Option Pricing When the Variance Changes Randomly: Theory, Esimation, and an Application", *Journal of Financial and Quantitative Analysis* 22, 419-438.

[80] Scott, L. (1991), "Random Variance Option pricing", *Advances in Futures and Options Research* Vol. 5, 113-135.

[81] Shackleton, N.J. and M.A. Hall (1989) "Stable Isotope History of Pleistocene at ODP Site 677",*Proceedings of the Ocean Drilling Program* 111, 295-315.

[82] Shephard, N. (1993), "Fitting Nonlinear Time Series Models With Applications to Stochastic Variance Models", *Journal of Applied Econometrics* 8, 135-152.

[83] Shephard, N. (1995), "Statistical aspects of ARCH and stochastic volatility", Discussion Paper No. 94, Nuffield College, Oxford.

[84] Smith, A.F.M. and Roberts, G. (1993), "Bayesian Computations via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods", *Journal of the Royal Statistical Society, series B* 55,3-23.

[85] Stein, E.M. and Stein, J.C. (1991), "Stock Price Distributions with Stochastic volatility: An Analytic Approach", *Review of Financial Studies* 4, 727-752.

[86] Taylor, T. and Wong, W. (1987), "The Calculation of Posterior Distributions By Data Augmentation. *Journal of American Statistical Association* 82, 528-549.

[87] Taylor, S.J. (1994), "Modelling Stochastic Volatility: A Review and comparative Study. *Mathematical Finance* 4, 183-204.

[88] Thomson P.J. and Robinson, P.M. (1996), "Estimation of Second-Order Properties from Jittered Time Series", *Annals of Institute of Statistics and Mathematics* 48, No. 1, 29-48.

[89] Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions (With discussion)", *Annals of Statistics* 22, 1701-1762.

[90] Titterington, D.M., Smith, A.F.M. and Makov, U.E. (1985), "Statistical Analysis of Finite Mixture Distributions", Chichester, Wiley, New York.

[91] Uhlig, H. (1993), "Bayesian Vector Autoregressions With Stochastic Volatility", Unpublished draft, Princeton Unversity, Dept. of Economics.

[92] Vetzal, K. (1992), "Stochastic Short Rate Volatility and the Pricing of Bonds and Bond Options", Unpublished Ph.D. thesis, University of Toronto.

[93] Watanabe, T. (1992), "A Nonlinear Filtering Technique for Estimating Stochastic Volatility Models With an Application to Daily Returns in the New York Stock Exchange", working paper, Yale University, Dept. of Economics.

[94] West, M. (1993), "Approximating Posterior Distributions by Mixtures", *Journal of the Royal Statistical Society* (Series B), 54, 553-568.

[95] West, M. (1994), "Elements of Spectral Analysis", (Lecture Notes on Time Series) ISDS Discussion Paper 94-29, Duke University.

[96] West, M. (1995), "Bayesian Inference in Cyclical Component dynamic linear models", *Journal of the American Statistical Association* 90, 1301-1312.

[97] West, M. (1996), "Bayesian Time Series: Models and Computations for the Analysis of Time Series in the Physical Sciences", *Maximum Entropy and Bayesian Methods 15,* (eds: K. Hanson and R. Silver), Kluwer (pp 23-34).

[98] West, M. and Prado, R. (1997) "Exploratory Modelling of Multiple Non-Stationary Time Series: Latent Process Structure and Decompositions", In Modelling Longitudinal and Spatially Correlated Data, (ed: T. Gregoire), Springer-Verlag (forthcoming).

[99] West, M. (1996), "Some Statistical Issues in Palaeoclimatology" (with discussion), Bayesian Statistics 5, (eds: J.O. Berger, J.M. Bernardo, A.P. Dawid and A.F.M. Smith), Oxford University Press.

[100] West, M. (1997), "Time Series Decomposition", *Biometrika* 84, (in press).

[101] West, M. and Harrison, J. (1997), Bayesian Forecasting and Dynamic Linear Models. (2nd Ed.) Springer-Verlag, New York.

[102] Wiggins, J. (1987), "Option Values Under Stochastic Volatility: Theory and Empirical Evidence", *Journal of Financial Economics* 19, 351-372.

Biography

Yang Chen was born in Beijing, P.R. China, on July 23, 1970. He received his B.S. in Mathematics and Computer Science from Central Missouri State University in Warrensburg, Missouri, in May 1993, graduating summa cum laude, Phi Beta Phi. He received his M.S. in Statistics from Duke University in Durham, North Carolina, in May 1995.

During his undergraduate study at Central Missouri State University, he received the Central Missouri State University Distinguished Student Award, the Mathematics Department Distinguished Student Fellowship and the Phi Beta Phi Graduate Study Fellowship. During his four-year study at Duke University, he received one semester graduate fellowship and three and a half year teaching assistantship. During the three summers of his Ph.D study, he participated in summer internship programs at the National Institute of Statistical Sciences in Research Triangle Park, North Carolina and J.P. Morgan Company in New York. His work at J.P. Morgan resulted in the first part of his dissertation. He will be joining the Corporate Risk Management Group at Prudential Securities in New York upon graduation.