The Pennsylvania State University
The Graduate School

BAYESIAN ANALYSIS OF MULTIVARIATE REGIME SWITCHING COVARIANCE MODEL

A Dissertation in
Statistics
by
Lu Zhang

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2010

The dissertation of Lu Zhang was reviewed and approved* by the following:

John C. Liechty
Professor of Marketing and Statistics
Dissertation Advisor, Chair of Committee

Runze Li
Professor of Statistics
Chair of Graduate Program

Murali Haran
Professor of Statistics

Timothy T. Simin
Professor of Finance

James Rosenberger
Professor of Statistics
Acting Department Head

*Signatures are on file in the Graduate School.

Abstract

We propose a new multivariate regime switching covariance model, where the covariances are decomposed into volatilities and correlations, both of which are regime switching. The model specifies an independent regime switching process for the volatilities of each asset, and one process for the correlation matrix. It is the first time that the volatility and correlation regimes are modeled simultaneously. From an in-sample perspective, it helps identify the relationship between the volatility and correlation dynamics. From a forward looking perspective, this model can potentially make good forecast of financial crisis where the market enters a high volatility and high correlation regime at the same time. Our model, along with the proposed Markov Chain Monte Carlo (MCMC) methods, contributes to solving three important technical issues. First, we model the unobserved regime switching process by a jump chain and waiting times between jumps. We can use both Exponential and Gamma distribution to describe the waiting time. This specification allows us to generalize the hidden regime process to be non-Markovian, which provides a better fit for empirical data that have seasonal switches in volatility levels. Secondly, we use a shrinkage model for the off-diagonal elements of the correlation matrix, which imposes an average correlation on each regime. This allows us to clearly represent and identify the latent correlation regimes. Third, since missing data is a challenge in real data analysis, we introduce a Bayesian imputation method which can accurately recover missing values, which can occur for example over different holidays for indices from different countries.

Based on the structure of our model, we also introduce a portfolio allocation strategy where a portfolio manager re-balances portfolio weights whenever a switch in regime is detected. Such a strategy keeps a good balance between stock return and risk, and at the same time saves portfolio adjustment cost. We discuss examples on simulated data set, natural gas commodity data and weekday international market indices.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I am most grateful and indebted to my thesis advisor, Dr. John C. Liechty, who has been my PhD advisor during my PhD study at Pennsylvania State University. As a joint faculty of Marketing and Statistics, he brought to me many inspiring ideas which aroused my interests in business statistics problems. He is an excellent Bayesian expert, giving me critical guidance and suggestions for my research. During the past three year working with him, he spent a lot of time and effort educating and helping me: we meet every week to discuss recent research progress; and he is responsive whenever I need help. He is always patient and supportive. He treats me as a collaborator, sharing his personal experiences and his understanding of interesting academic topics with me, and allows me a lot of freedom in exploring new ideas.

It also my great pleasure to convey my gratitude to Dr. Runze Li for being so considerate and supportive. He is always there to provide advice and help whenever I am confused about academic, life or future, so he is like an enlightening friend. I also want to thank him for his generous financial support which allows me to focus on my research and job hunting.

I would like to thank Dr. Murali Haran. I took two most important courses with him, MCMC and Spatial model, which are extremely useful in my research. His suggestions and comments on my work are very important and encouraging. I am also indebted to Dr. Tim Simin who is a co-author of my paper. As a finance faculty, he provides constructive insights which lead to very interesting applications of our statistical model to finance problems. I am so lucky and grateful to have Professor Runze Li, Murali Haran and Tim Simin in my committee.

I will also gratefully thank my department, statistics department in Penn State University. We have first class faculties and peer students. I feel lucky to receive my Phd education with this department. I have taken courses with Dr. Bruce Lindsay, Dr. Naomi Altman, Dr. Tom Hettmansperger, Dr. Murali Haran, Dr. David Hunter, Dr. Steve Arnold, Dr. Jogesh Babu, Dr. Francesca Chiaromonte, Dr. John Fricks, Dr. Bing Li, Dr. Joseph Schafer. Special thanks to Dr.Francesca Chiaromonte who was my temporary advisor in the first year and helped me get accustomed to graduate life.

Besides the academic achievements, the department also left in my mind a lot of touching moments. Everyone in this department is like a family member who is willing to extend help whenever in need. I have two examples. Dr. Altman, although very busy, spent her spare time helping me rehearse my job talk and provide critical suggestions. Dr. Rosenberger, as an online course coordinator, assigned me online course grader TAship which allows me to re-unite with my husband. Words fail me to express my appreciations to them.

I would thank my all my friends, who shared my happiness and bitterness, x enjoyed life with me, and exchanged wisdom with me.

I am indebt to my dear husband Yang Song, who has been an indispensable part of my life and research. His endless love to me and our family has been the greatest momentum for me to finish my PhD study at Penn State. As a PhD in CSE department at Penn State and a researcher at Microsoft Research, he also devoted his time and effort to help me on my research leveraging his expertise in machine learning and computing skills. Yang and I have worked together on several interesting and challenging interdisciplinary problems. Our research results have been published in top-tier computer science conferences and journals.

Finally, I would like to thank my parents, Jiansheng Zhang & Sujun Zeng, who raised me and supported me throughout for all my 26 years' life. They are always there, listening to me, worrying for me, happy for me, and they have devoted all their lives to me.

Dedication

This thesis is dedicated to my dearest parents, Jiansheng Zhang & Sujun Zeng, and my beloved husband, Yang Song.

Chapter 1

Introduction

There is a large body of empirical work establishing that the covariances of financial assets vary over time. Accurately modeling the covariance dynamics is crucial in applications on option pricing, risk management and portfolio allocation. The famous Black-Scholes option pricing model is based on a constant variance assumption. However, the performances of the model can be greatly improved if we generalize the model to allow for dynamic volatilities. In the risk management area, Value-at-Risk (VaR) is a widely used measure of risk for a specific portfolio of financial assets. VaR predicts the loss in a worst-case scenario based on the forecasted distribution of financial assets. To avoid or to be prepared for any unaffordable losses in the future, it is important to have a model that can accurately forecast the distribution of financial assets' prices, particularly the correlation between asset prices. With respect to portfolio allocation, portfolio managers normally adjust or rebalance the proportions of financial assets in a portfolio monthly, quarterly or every half a year. The purpose is to adapt a portfolio to maximize its return and minimize risk based on up-to-date information.

The model in Bollerslev [8] is one of the most popular multivariate time varying covariance model. In this model, the covariances of a portfolio of asset returns are decomposed into standard deviations (volatilities) and correlations. The volatilities are modeled as time varying while the correlation is hypothesized to be constant. Based on such a decomposition, the covariance matrix is estimated in two steps: first individual volatility processes are estimated and then conditional on the volatilities, the correlation matrix is estimated. There is a large number

of competing methods in the literature on how to model the univariate standard deviations[1]. In addition to the univariate models of time varying volatility, the constant correlation assumption has also been extended to be time varying[2].

An interesting branch of the time varying models are the regime switching models. These type of models are not new in econometrics research. The objective of introducing regime switching is to allow for abrupt structure changes in the model. For example, in financial market, stock prices may follow a diffusive process with distinctively different parameters during a bull or a bear market. As a result, when a market suddenly becomes bear from bull, a diffusive model is not enough to explain the fundamental change and can mis-specify the true underlying mechanism. One example is that the volatility persistence is spuriously high when no regime switching is incorporated in the volatility modeling. While the persistence estimates drop significantly after regime switching stochastic volatility models are suggested [33][3]. As for the correlation, Pelletier [44] introduces a Markov switching correlation model where the correlation matrix is constant within a regime, but changes when each regime changes. This model reconciles the Dynamic Conditional Correlation (DCC)-GARCH [17] and Constant Conditional Correlation (CCC)-GARCH model [8]. On one hand it explains the dynamic nature of the correlation matrices, and on the other hand, it is easy to impose that the covariance matrix is PSD.

However, to our knowledge, simultaneously modeling of regime switches in both the volatilities and the correlations has never been implemented because estimating a large number of switching parameters raises the concern of model instability [44]. In practice, some researchers only assume a regime switching process for the volatilities of one major asset in a portfolio and then calculate correlations according to

[1]Robert F. Engle[16] proposed an Autoregressive Heteroscedastic (ARCH) model for time varying volatilities, which was generalized [7] to the GARCH model. In parallel to the ARCH/GARCH approaches, stochastic volatility models were introduced [28] and are gaining popularity nowadays.

[2]Existing models include the dynamic conditional correlation (DCC)-GARCH model[17] where parameters in the correlation matrix follow a GARCH type autoregressive relation. A counterpart to this model is DCC-MSV model [1] where the stochastic volatility formula replaces GARCH for correlation matrices.

[3]Hamilton [22] provided a switching ARCH model (SWARCH) where ARCH specifications accommodated regime switching. Analogously, Markov switching stochastic volatility models are also found in the literature[49][48][27].

the corresponding volatility regimes[12]. Other authors assume that correlation matrix determines the hidden regime dynamics while standard deviations follow a completely different mechanism which is modeled separately. Our goal in this study is to introduce a model and its related Bayesian inference methodology to address the simultaneous modeling problem. We resort to the MCMC sampler where in each sweep of update, we sample univariate volatilities for each asset and their correlations conditional on the remaining parameters. This framework finds application in multiple areas. Looking backward, we can use it to identify structural switches associated with historic events. We can also use it to find the relationship between volatility regimes and correlation regimes, which is helpful in explaining econometric phenomena. Looking forward, we can forecast financial crisis which is characterized by the simultaneous occurrence of a high volatility and high correlation regime.

In the implementation of our general regime switching covariance model framework, we address several technical implementation challenges. First, we propose a regime switching stochastic volatility model to fit univariate standard deviations where the regime process can be non-Markovian. The difficulty with inference comes from the two layers of latent processes in this model: the hidden volatility process $\mathbf{h}$ and the hidden regime process $\mathbf{D}$. We use the Kalman filter based [49][32] method to infer the unobserved volatilities. As for the regime process $\mathbf{D}$, most existing regime switching models assume a discrete time Markovian structure of the switching process with an associated transition matrix. We break from this approach and leverage a continuous-time model of $\mathbf{D}$ which is assumed to consist of jumps and waiting times [36]. In [36], the waiting times follow an exponential distribution. Because of the memoryless property of an exponential distribution, the process $\mathbf{D}$ is Markovian. In our study, we allow the waiting time to be a gamma distribution, so that the process $\mathbf{D}$ is not restricted to be Markovian. The introduction of two parameters in a gamma distribution provides a richer structure for $\mathbf{D}$. Since the gamma density is more concentrated compared to the exponential density, our proposed model is especially useful for data that have somewhat regular switches in regimes and hence consistent length of waiting times. Following [36], we use a reversible jump MCMC method to make inference for the regime process. The number of regimes is determined completely by data. We decide

on the best mode by considering several model comparison tools. Application to natural gas commodity data demonstrates the value of this extension for some cases.

Another extension of our univariate model is its flexibility in accommodating not only different mean volatility levels, but also different persistence and variance of volatilities across regimes. Empirical analysis on JPMorgan stock data shows that a high volatility regime is associated with high persistence and low variance of volatility.

In addition to the time varying volatilities, we follow Pelletier [44] and model correlations with regime switches: that is, the correlation matrix is assumed to be constant within a regime, but changes when each regime changes. Due to the positive semi-definite (PSD) constraints, there are two limitations in existing correlation models: the first one is in the GARCH type modeling of dynamic correlations [17]. A step of re-scaling the correlation diagonals to one introduces non-linearity; the second one is in the work of Pelletier [44], where the model cannot accommodate a correlation matrix which has negative elements. We improve upon these models by adopting a Bayesian hierarchical modeling of correlation matrix [35]. The off-diagonal elements of a correlation matrix are assumed to shrink towards a common mean. With the common mean representing a correlation matrix, it is easier to identify latent correlation regimes. Our simulations indicate that this estimation method successfully recovers all the parameters, as well as the hidden regime processes in both the univariate and multivariate stochastic volatility framework.

We apply our regime switching covariance model to a portfolio of daily international indices data. When we line up the indices data by calender days, we encounter the problem of a large number of missing data due to different holiday schedules in different countries. We solve this problem using a Bayesian imputation strategy which extracts information from the estimated volatility and correlation structures and makes inferences of the missing values. Our strategy improves parameter estimation compared to other imputation methods.

Besides the technical improvements mentioned above, we demonstrate the use of our model on portfolio allocation problem. A portfolio manager determines the proportion of each asset in a portfolio with the goal of increasing return and

reducing risk. In practice a portfolio manager adjusts or rebalances the asset weights based on recent data and information. Typically a portfolio is re-balanced on a fixed schedule, e.g., monthly or quarterly. There is little justification for these timing strategies and we observe that it is a waste of money to re-balance a portfolio when there is not much change in the structure of asset values. One alternative would be to re-balance a portfolio only when a regime switch in either the volatility or the correlation is detected. Both simulation and empirical analysis show that our strategy saves re-balance cost and increases average Sharpe Ratios compared to regularly timed portfolio re-balancings.

The rest of the thesis is organized as follows. We focus our discussion on the univariate volatility models in Chapter 2 and Chapter 3. In Chapter 2, we review the univariate stochastic volatility model and its extensions in the literature. In Chapter 3, we present the technical details and applications of our univariate regime switching stochastic volatility model. This includes the inferences for the hidden Markov/non-Markov processes by the reversible jump MCMC method and the parallel tempering strategy to improve mixing. Simulation results demonstrate the ability of our model to fit the data and also help to identify the value of our model selection criteria. We work with real data with both a Markovian and a non-Markovian regime process. The remaining chapters are devoted to multivariate models. In Chapter 4, we review the multivariate covariance modeling literature. In Chapter 5, we provide the detailed multivariate regime switching covariance model along with its inference methods. In Chapter 6, We introduce our Bayesian missing data imputation method and portfolio allocation strategy, and include some simulation results as well. In Chapter 7, we apply all our proposed techniques on daily international indices data. We conclude with a discussion of possible future research areas in Chapter 8.

Chapter 2

Review of Univariate Volatility Model

Time varying volatilities of financial asset returns are well documented. The autocorrelations of financial return time series tend to decay relatively fast while the autocorrelations of its second moment persist. Mandelbrot [39] observed volatility clustering phenomenon, where "large changes in the volatilities tend to be followed by large changes-of either sign-and small changes by small changes". In this chapter, we give a review of univariate volatility models in the literature. There are two main streams of research in this area: ARCH/GARCH models and stochastic volatility models. In addition to presenting the basic models, we review popular extensions.

2.1 Autoregressive Heteroscedastic (ARCH) Model

In 1982, the famous ARCH model was proposed by Engle [16], which has been commonly used to explain the volatility clustering characteristics. ARCH model is a discrete time model and is widely applied in situations where the volatility of a time series is of a major concern. It also finds applications in asset pricing, option pricing, asset allocation and risk management.

The basic ARCH(p) model can be written as:

$$\begin{aligned} Y_t - X_t\beta &= \sqrt{h_t}Z_t; \\ h_t &= a_0 + \sum_{i=1}^{p} a_i(Y_{t-i} - X_{t-i}\beta)^2. \end{aligned} \tag{2.1}$$

where $\{Y_t\}$ denotes the log-returns. $X_t\beta$ is the mean for Y_t and h_t is the volatility. Z_t is assumed to be i.i.d Normal(0,1). The process h_t is a function of past squared residuals (or mean corrected log returns). The parameter a_0 is greater then 0 and a_i are assumed to be nonnegative parameters because h_t is a variance.

Empirical experience calls for a large p in the conditional variances formula 2.1 because h_t has high persistency. This leads to a generalization of the ARCH model, namely GARCH(p,q), which was proposed by Bollerslev [7] as shown in Equation 2.2. In the GARCH(p,q) model, the volatility h_t is a function of past squared residuals $(Y_{t-i} - X_{t-i}\beta)^2$ and previous volatilities h_{t-i}. Hansen and Lunde [23], compared 330 candidate models and concluded that no model can outperform GARCH(1,1) when fitted to daily exchange rate data.

$$\begin{aligned} Y_t - X_t\beta &= \sqrt{h_t}Z_t; \\ h_t &= a_0 + \sum_{i=1}^{p} a_i(Y_{t-i} - X_{t-i}\beta)^2 + \sum_{i=1}^{q} b_i h_{t-i}. \end{aligned} \tag{2.2}$$

It is interesting to note that with a GARCH(1,1) model, the conditional volatility depends on the whole path of return and the impact of shocks to return stays for long. This phenomenon is evident when you reframe the GARCH model by substituting h_{t-i} recursively in the following way:

$$h_t = a_0 + \sum_{i=1}^{\infty} \phi_i(Y_{t-i} - X_{t-i}\beta)^2. \tag{2.3}$$

As with the ARMA models, the GARCH extension provides a parsimonious model which is easier to estimate in practice.

2.2 Generalization of ARCH/GARCH model

The standard ARCH/GARCH models make several assumptions which are not true in practice. For example, the autocorrelations of the squared returns are assumed to decay at an exponential rate. However, empirical evidence has frequently suggested much greater degree of persistence in the autocorrelation, or a "long memory" in the square of returns. Moreover, empirical work suggests that there are two factors determining the return and volatility processes: A diffusive, that is, slowly-changing factor and a rapidly changing factor. This rapidly changing factor is named a "jump". The jump effect is not taken into consideration in standard ARCH/GARCH models.

Various extensions of the basic ARCH/GARCH model are proposed with different assumptions and motivations. We introduce several generalizations which are relevant to our research.

2.2.1 ARCH Model With Fat Tails

A standard GARCH model cannot fully explain high kurtosis, heavy tails and extreme events which happen in reality. Therefore, Bollerslev [7] introduced a GARCH model with t-distributed innovations, which improved but did not completely solve this problem [46]. Nelson [41] proposed to use a generalized error distribution and Engle and Gonzalez-Rivera [18] applied a non-parametric approach. Other authors, including Bauwens [4] and Bai [3] proposed modeling the innovations distribution with a mixture of two zero mean normal distributions with different variances. Such models can capture volatility clustering, high kurtosis, heavy tails and the presence of extreme events.

2.2.2 ARCH Model With Leverage Effect

Leverage effect is a common phenomenon in financial returns. The usual claim is that when there is bad news, which decreases the price (return), it makes the firm riskier by increasing future expected volatility and vice versa. In a basic GARCH model, because h_t cannot be negative, it is modeled as a linear combination of squared terms (with nonnegative weights). It cannot distinguish whether the re-

turn shock term Z_t is positive or negative. As a result this symmetry feature cannot explain the asymmetric leverage effect [41].

A generalization of the ARCH/GARCH model to capture the leverage effect is called EGARCH model. In this model, $\log(h_t)$ substitutes h_t and it is modeled as a linear function of time and lagged Z_t. In this way the nonnegative constraint can be avoided.

The general form of the volatility process is:

$$\log(h_t) = \alpha_0 + \sum_{i=1}^{q} \beta_j \log(h_{t-i}) + \sum_{i=1}^{p} g(Z_{t-i}). \tag{2.4}$$

To accommodate the asymmetric relation between return and volatility, both sign and magnitude of Z_t should be considered. One choice of $g(Z)$ can be

$$g(Z_t) = \theta Z_t + \gamma[|Z_t| - E(Z_t)]. \tag{2.5}$$

where θ and γ are constant parameters.

2.2.3 ARCH model with Switching Regimes

The financial market is sometimes quite calm while some other times it is highly volatile. ARCH/GARCH models (including some of their extensions) cannot account for swift shifts in market structures and they tend to perform poorly with respect to model fitting and forecasting. Moreover, basic ARCH/GARCH models imply a high degree of persistence in the volatility which may not exist in empirical data. Hamilton (1994)[22] proposed a switching ARCH model (SWARCH), which accomodates possible structural changes in the ARCH process.

The SWARCH model depends on **D**, an unobserved sequence whose value D_t can take on the values of 1,2,...,K, where K is the total number states. Each state represents the underlying volatility structure in financial market. **D** is assumed to be a discrete time Markov Chain with a transition matrix P_{K*K}.

A simple SWARCH model can be written as:

$$Y_t \;=\; X_t\beta + \sqrt{g_{D_t}}\sqrt{h_t}Z_t;$$

$$h_t = a_0 + \sum_{i=1}^{p} a_i Z_{t-i}^2 h_{t-i} + \sum_{i=1}^{q} b_i h_{t-i}. \quad (2.6)$$

where $\{h_t\}$ is the volatility process which can be modeled either using a standard ARCH/GARCH model or in any of its generalied forms. The residuals $\sqrt{h_t}Z_t$ in the log-return equation is multiplied by a constant g_{D_t}. When the process is in regime 1 denoted by $D_t = 1$, we multiply g_1, when $D_t = 2$, multiply g_2 and so on. The idea is to model the changes in regime as a change in the scale of the volatility process.

This model is called a K state qth order Markov switching ARCH process, and is denoted by SWARCH (K,q) when h_t is modeled as ARCH(q) process. Or it is called SWARCH-L(K,q) when leverage effect is considered to model h_t.

2.3 Stochastic Volatility Model Literature

An alternative to the ARCH/GARCH framework is the model in which the variance follows a latent stochastic process. This model is called stochastic volatility models which were introduced in 1994 by Jacquier [28] and Shephard [47].

The basic stochastic volatility model in discrete time is:

$$\begin{aligned} y_t &= \exp(h_t/2)\varepsilon_t \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \tau\eta_t, \\ h_1 &\sim N\left(\mu, \frac{\tau^2}{1-\phi^2}\right) \end{aligned} \quad (2.7)$$

where y_t is the mean corrected return, and h_t is the log volatility at time t which is assumed to follow a mean reverting first order autoregressive stationary process ($|\phi| < 1$). Compared to the ARCH/GARCH models, there are two shock terms, ε_t and η_t, for the asset return and the volatility respectively, which are uncorrelated standard normal white noise. The error ε_t is a transient shock because it only influences y_t; however η_t has a more persistent influence because it has an impact on h_t, and through the autoregressive structure of $\mathbf{h}$, such an impact is transferred to $\mathbf{h}^t = (h_{t+1}, ..., h_N)$. These two stochastic shocks in the stochastic volatility model provide more flexibility compared to ARCH/GARCH models, but at the

same time, they complicate estimation because **h** is now an unobserved, non-deterministic process.

The parameter μ in Equation 2.7 is the mean level of volatility, ϕ is the volatilities' mean reverting persistency parameter and τ^2 is the variance of the volatility process. We denote the parameters for this volatility equation by $\theta = (\mu, \phi, \tau)$.

2.4 Generalization of Stochastic Volatility model

Analogous to the ARCH/GARCH models, extensions of the basic stochastic volatility model have been proposed to explain more complicated financial phenomena.

2.4.1 Stochastic Volatility Model With Fat Tails

Nielsen, Nicolato and Shephard [43] addressed the fat tail problem by including a scale mixture variable λ_t where the realization of this variable is considered a latent variable.

The basic model is extended as:

$$\begin{aligned} y_t &= \exp(h_t/2)\sqrt{\lambda_t} z_t \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \tau\eta_t \\ (z_t, \eta_t) &\sim N(\mathbf{0}, \Sigma) \\ \lambda_t &\sim p(\lambda_t|\nu) \end{aligned} \tag{2.8}$$

where λ_t is assumed to follow i.i.d inverse gamma distributions, or that $\nu/\lambda_t \sim \chi^2_\nu$. This implies that $\varepsilon_t = \sqrt{\lambda_t} z_t$ follows a student $t(\nu)$ distribution.

2.4.2 Stochastic Volatility Model With Leverage Effect

A leverage effect can be incorporated by simply correlating the two error terms in the return and volatility processes [29]. According to the definition of a leverage effect, that is, an increase in return will decrease the volatility and vice versa, the correlation ρ between the two error terms should be negative.

Hence a stochastic volatility model with leverage effect is given by:

$$\begin{aligned} y_t &= \beta \exp(h_t/2)\varepsilon_t, \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \tau\eta_t, \ . \\ h_1 &\sim N(\mu, \tfrac{\tau^2}{1-\phi^2}), \end{aligned}$$

where the correlation between ε_t and η_t is ρ and $\rho < 0$.

2.4.3 Stochastic Volatility Model With Markov Switching

Lamoureux and Lastrapes [33] suggested that the documented high persistency of variance may have been overestimated because structural shifts in the market are not taken into account. Hamilton [22] included a Markov switching in the ARCH /GARCH models, and So [49] incorporated regime switching structures into the stochastic volatility models. In these models:

$$h_{t+1} = \mu_{D_{t+1}} + \phi(h_t - \mu_{D_{t+1}}) + \tau\eta_t \tag{2.9}$$

where

$$\mu_{D_{t+1}} = \gamma_1 + \sum_{j=2}^{K} \gamma_j I_{jt} \tag{2.10}$$

and I_{jt} is an indicator variable that equals 1 when $D_t \geq j$. The process $\{D_t\}$ is an unobserved discrete time regime process which is a K state first order Markov Process, or Markov Chain. The idea is to model the changes in regime as a change in the mean level of the volatility process. This model is denoted as MSSV(K), and Bayesian methods are applied for inference.

2.4.4 Markov-Switching Diffusion Model

Smith [48] proposed a Markov switching diffusion model and compared it to the MSSV(K) model on short-term interest rate. Markov switching diffusion model differs from the MSSV(K) model in that the volatilities are constant within each regime. In our representation, the Markov switching diffusion model is:

$$y_t = \sigma_{D_t}\varepsilon_t \tag{2.11}$$

where D_t represents the particular regime at time t and ε_t is white noise. The parameter σ_i is the constant standard deviation of y_t in regime i. Quasi maximum likelihood estimation was implemented and the author concluded that "Markov-switching diffusion or stochastic volatility, but not both, are needed to adequately fit the data" and the "Markov-switching diffusion model is the best in terms of forecasting".

2.4.5 Stochastic Volatility Models with Markov Regime Switching State Equations

Pereira [27] generalized the MSSV(K) model and introduced a stochastic volatility model with Markov Regime Switching State Equations(SVMRS). This model is in essence a stochastic volatility model, but it allows μ, ϕ and τ to change by regimes, which is more general than the MSSV(K) model where only μ differs by regimes.

$$h_{t+1} = \mu_{D_{t+1}} + \phi_{D_{t+1}}(h_t - \mu_{D_t}) + \sigma_{D_{t+1}}\eta_t \tag{2.12}$$

As in the MSSV(K) model, D_t is an unknown discrete regime indicator variable which is Markovian, and Quasi maximum likelihood estimation is adopted in parameter inferences.

2.5 Inference for the Hidden Regime Process

While fat tails, leverage effect, jumps and long memory properties are important in modeling volatilities, we leave them for future discussion and research. In this dissertation, we focus on modeling regime switches in multivariate stochastic volatility models.

The main challenge in making inference for stochastic volatility models lies in the inference for the unobserved stochastic process $\mathbf{h}$. A variety of algorithms have been proposed to make inference for $\mathbf{h}$, among which a filtering based algorithm [32] has proved to be efficient and easy to implement. We will describe this algorithm in detail in Chapter 3.

Introducing a regime switching process $\mathbf{D}$ into the stochastic volatility model

adds another layer of complexity, which makes model estimation even more complicated. As a result efficient inference methods for the hidden Markov chains are desirable. All existing algorithms to infer $\mathbf{D}$ are based on "filtering". One reason is that there is an abundant literature and softwares at hand which can implement filtering directly. In this section, I introduce several popular filtering based approaches for the inference of $\mathbf{D}$.

2.5.1 Quasi Maximum Likelihood (QML) Based Method

Pereira [27] proposed a Quasi Maximum likelihood method through a Kalman filter to estimate $\mathbf{D}$ in his SVMRS model.

In the SVMRS model, the observational equation $y_t = \exp(h_t/2)\varepsilon_t$ is transformed to a linear model $y_t^* = \log y_t^2 = h_t + \psi_t$, where ψ_t is a $\log \chi^2$ distribution. Following Harvey, Ruiz and Shephard [24], ψ_t is treated as though it were from $N(0, \frac{\pi^2}{2})$, and parameters are estimated by maximizing the resulting quasi-likelihood function.

According to the SVMRS model, given the knowledge of the parameters $(\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma})$, the stochastic volatility equation is determined by the pair (D_{t+1}, D_t). Because only two possible regimes are assumed in their model, there are four combinations of (D_{t+1}, D_t):

$$\begin{aligned}
&D_{t+1} = 0, D_t = 0, \quad then \quad h_{t+1} = \mu_0 + \phi_0(h_t - \mu_0) + \tau_0\eta_t \\
&D_{t+1} = 0, D_t = 1, \quad then \quad h_{t+1} = \mu_0 + \phi_0(h_t - \mu_1) + \tau_0\eta_t \\
&D_{t+1} = 1, D_t = 0, \quad then \quad h_{t+1} = \mu_1 + \phi_1(h_t - \mu_0) + \tau_1\eta_t \\
&D_{t+1} = 1, D_t = 1, \quad then \quad h_{t+1} = \mu_1 + \phi_1(h_t - \mu_1) + \tau_1\eta_t
\end{aligned} \tag{2.13}$$

The number of combinations increases quadratically with the increased number of possible regimes. Iterating from $t = 0$ through a Kalman filter which is detailed in the appendix, for $D_{t-1} = i$ and $D_t = j$, one can calculate

$$v_t^{ij} = y_t^* - h_{t|t-1}^{ij} \tag{2.14}$$

and

$$f_t^{ij} = E[(h_t - h_{t|t-1}^{ij})^2 | D_t = i, D_{t-1} = j, I_{t-1}] + \sigma_\phi^2 \tag{2.15}$$

where v_t^{ij} and f_t^{ij} are the updated residual and variance respectively. The symbol I_t is the information set up to time t. QML assumes normality for y_t^*, therefore when $D_{t-1} = i$ and $D_t = j$,

$$f(y_t^*|D_t = i, D_{t-1} = j, I_{t-1}) = \frac{1}{\sqrt{2\pi f_t^{ij}}} \exp \frac{-(v_t^{ij})^2}{2f_t^{ij}} \tag{2.16}$$

.

When the conditional probability $p(D_t = i, D_{t-1} = j|I_{t-1})$ is updated iteratively by a Kalman filter, the likelihood for this model is:

$$\boldsymbol{L}(\mathbf{y}^*|\boldsymbol{\theta}) = \sum_{t=1}^{T} \sum_{ij} \log[f(y_t^*|D_t = i, D_{t-1} = j, I_{t-1}) p(D_t = i, D_{t-1} = j|I_{t-1})] \tag{2.17}$$

which can be easily calculated. Numeric maximization of the loglikelihood function leads to the QML estimates of $\boldsymbol{\theta}$. Various starting values of $\boldsymbol{\theta}$ must be tried in order to avoid local maximization. A typical smoothing technique is applied to extract the smoothed information $h_t|I_T$ and $s_t|I_N$ where N is the length of data. Detailed description of this approach can be found in Pereira [27].

2.5.2 Bayesian Method

The Markov switching stochastic volatility models [49] are also estimated using Bayesian MCMC methods. The complication with this model rests in the second layer of the model with the latent Markov chain $\mathbf{D}$. In the paper, So [49] adopted a multimove sampler to simulate $\mathbf{D}$ jointly from its full conditional distribution.

A decomposition of the full conditional distribution of $\mathbf{D}$ leads to

$$f(\mathbf{D}|\mathbf{y}, \mathbf{h}, \boldsymbol{\theta}) = f(D_N|\mathbf{y}, \mathbf{h}, \boldsymbol{\theta}) \prod_{t=1}^{N-1} f(D_t|\mathbf{y}, \mathbf{h}, \boldsymbol{\theta}, \mathbf{D}^{t+1}), \tag{2.18}$$

where $\mathbf{D}^t = (D_t, ..., D_N)$. As a result one can sample $\mathbf{D}$ jointly if they know how to sample from $f(D_N|\mathbf{y}, \mathbf{h}, \boldsymbol{\theta})$ and $f(D_t|\mathbf{y}, \mathbf{h}, \boldsymbol{\theta}, \mathbf{D}^{t+1})$ respectively.

Using the Markovian property and Bayes rule ,

$$f(D_t|\mathbf{y}, \mathbf{h}, \boldsymbol{\theta}, \mathbf{D}^{t+1}) \propto f(D_{t+1}|D_t) f(D_t|y_t, h_t, \boldsymbol{\theta}) \tag{2.19}$$

where $f(D_{t+1}|D_t)$ is given by the transition matrix of $\mathbf{D}$. A discrete filter developed by Carter and Kohn [10] can be applied to evaluate $f(D_t|y_t, h_t, \boldsymbol{\theta})$, which is detailed in Section in 3.5.1. As a result, a simultaneous sampling of $\mathbf{D}$ is straightforward.

2.6 Motivation for Proposed Univariate Model

The regime switching stochastic volatility models described above can be improved in several aspects.

First, all of the existing regime switching models are based on a Markovian transition assumption for the regime process $\mathbf{D}$. While this assumption is valid in certain cases, we should be cautious about its exceptions. The determination of regime at time t may depend on D_{t^*}, where $t^* < t$ and t^* is random, which would make $\mathbf{D}$ non-Markovian.

Second, existing models are not flexible enough. In the MSSV(K) model [49], only the volatility mean level μ varies by regimes. In the Markov Switching model [48], volatilities are assumed to be constant within each regime. In the SVMRS model [27], all volatility parameters μ, ϕ, τ^2 are allowed to differ across regimes, however because of computational difficulties, only two regimes are considered in their framework. Empirical experience calls for a much more flexible model because there are probably more than two regimes in the market, and the mean reverting persistency and variance parameters should not be constrained to be identical under different circumstances.

Based on these observations, we propose a continuous-time regime switching stochastic volatility model, where the total number of regimes K is not restricted to be 2, where (μ, ϕ, τ^2) can differ across regimes, and where the hidden regime process $\mathbf{D}$ can be non-Markovian. We estimate parameters in a Bayesian framework and Gibbs sampler, slice sampler, reversible jump sampler all play roles in our inference. In addition to these standard sampling schemes, we use an offset mixture representation filter is applied to update the volatility process $\mathbf{h}$.

Chapter 3

Regime Switching Univariate Stochastic Volatility Model

Assume there are K volatility regimes. The canonical regime switching stochastic volatility model is:

$$\begin{aligned} y_t &= exp(h_t/2)\varepsilon_t, \quad \varepsilon_t \sim N(0,1) \\ h_t &= \mu_{D_t} + \phi_{D_t}(h_{t-1} - \mu_{D_t}) + \tau_{D_t}\eta_t, \quad \eta_t \sim N(0,1) \end{aligned} \tag{3.1}$$

where y_t is the mean corrected response variable, and h_t is the unobserved log volatility at time t which is assumed to follow a stationary first order autoregressive process ($|\phi| < 1$). The parameters (μ, ϕ, τ) are the volatility process parameters, indicating respectively the mean volatility level, the volatility mean reverting persistency, and the variance of the volatilities. $\mathbf{D} = \{D_t\}$ is the hidden regime indicator at time t. The errors ε_t and η_t are uncorrelated standard normal variables. As there are different sets of parameters, we denote $\boldsymbol{\theta} = (\boldsymbol{\mu} = (\mu_1, ..., \mu_K), \boldsymbol{\phi} = (\phi_1, ..., \phi_K), \boldsymbol{\tau} = (\tau_1, ..., \tau_K))$.

3.1 Non-Markovian Regime Process

Without exception, the hidden regime process $\mathbf{D}$ in existing regime switching volatility models is treated as a discrete time Markov Chain. Such a stipulation is necessary in order to apply the classical filtering methods in Markov switching

models [22]. We propose to adopt a continuous-time modeling of the regime process **D** [36] which can be generalized to allow for a non-Markovian structure of **D**.

In our framework, **D** is modeled in terms of a jump chain $(i_0, i_1, ...)$ and waiting times between jumps $(t_0, t_1, ...)$, which we call intervals. The waiting times follow a gamma $G(\alpha_k, \beta_k)$ distribution, $k = 1, ..., K$, where the parameters (α_k, β_k) can differ across different regimes. The jump transition matrix is $\mathbf{P} = \{p_{ij}\}_{K*K}$, where p_{ij} is the probability that **D** switches from state i to state j, given a jump occurs.

A special case of our model is the one in [36] where $\alpha = 1$, that is, the waiting time follows an exponential distribution. In this case, the hidden process **D** is equivalent to a discrete time Markovian switching process where the transition probability $\mathbf{Q} = \{Q_{ij}\}_{K*K}$ is given by

$$\begin{aligned} Q_{ij} &= p_{ij} * \lambda_i, \quad i \neq j, \\ &= 1 - \lambda_i, \quad i = j, \end{aligned} \tag{3.2}$$

where λ_i is the exponential distribution parameter for the ith regime.

When the waiting time follows a gamma distribution, it does not have the memoryless property of an exponential distribution, and hence the process is not Markovian anymore. Moreover the two parameters, shape and scale in a gamma distribution, makes the gamma density function more concentrated around certain values. This generalization may be more suitable for data with regular switches or consistent length of waiting times, i.e. data with a strong seasonal component. Our framework allows us to test this hypothesis in the following sections.

3.2 Flexibility of Model Parameters

At the beginning of our study, we were interested in determining which of the volatility parameter, μ, ϕ, *or* τ^2, are the underlying driving force that distinguishes volatility regimes. The most natural answer is that regimes should be identified by the volatility level μ. That is, regimes differ because they have different volatility levels, not because they have different persistence and variance parameters. We were also interested to explore whether two different regimes can share the same

volatility level, but differ only in the persistence or variance parameters of the volatilities. We experimented with simulation and real data and determined that different regimes are determined by different levels of volatility μ, rather than different persistence ϕ or variances of volatility τ^2. Therefore we proceed with our study assuming the μ drives and identifies the different regimes. To avoid problems with identification, we specify a certain order for the parameters $\boldsymbol{\mu}$ based on the states of the underlying chain, e.g., $\mu_1 < \mu_2 < ... < \mu_K$. We call this "Anchoring on $\boldsymbol{\mu}$". Our model is very flexible in that when $\boldsymbol{\mu}$ is anchored, ϕ or τ^2 can be the same or different across different regimes, which is decided by model comparison tools.

3.3 Likelihood Function

The marginal likelihood function $f(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{P}|\mathbf{y})$ is

$$f(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{P}|\mathbf{y}) \propto \int f(\boldsymbol{\theta}, \mathbf{h}, \mathbf{D}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{P}|\mathbf{y}) d\mathbf{h} d\mathbf{D} \tag{3.3}$$

This likelihood function is not closed form and difficult to integrate. Therefore we use a parameter augmentation strategy as in [32] where the augmented parameters in our model is

$$\boldsymbol{\Theta} = (\boldsymbol{\theta}, \mathbf{h}, \mathbf{D}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{P}) \tag{3.4}$$

We describe the MCMC algorithms used to update the augmented parameters for our model in the following sections.

3.4 Inference for the Regime Switching Stochastic Volatility Models

The posterior likelihood function for the augmented parameters is decomposed in Equation 3.5:

$$f(\boldsymbol{\theta}, \mathbf{h}, \mathbf{D}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{P}|\mathbf{y}) \propto f(\mathbf{y}|\mathbf{h}) f(\mathbf{h}|\boldsymbol{\theta}, \mathbf{D}) f(\mathbf{D}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{P}) f(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{P}) \tag{3.5}$$

We divide the augmented parameter space into five blocks: $\mathbf{h}$, $\mathbf{D}$, $\boldsymbol{\theta}$, $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $\mathbf{P}$. MCMC sampling algorithm is applied to generate samples for each block from its full conditional distribution.

We describe the inferences for each block of parameters separately in the following subsections and conclude with a general algorithm with an initialization strategy.

3.4.1 Inference for the Volatility Parameters

Conditional on the hidden regime process $\mathbf{D}$, we can estimate the volatility parameters for each regime as in a simple univariate stochastic volatility model. We follow the MCMC algorithms proposed in [32] throughout the paper. The only difference is that we order $\boldsymbol{\mu}$ in order to avoid parameter identification problems. This leads us to use a slice sampling method for $\boldsymbol{\mu}$. We illustrate the estimation of the volatility parameters for one regime, and it is easy to extend for all the other regimes.

Inference for $\boldsymbol{\tau^2}$

A conjugate prior $IG(\sigma_r/2, S_\sigma/2)$ is placed on $\tau_k^2,\ k = 1, ..., K$. The posterior distribution for τ_k^2 is:

$$f(\tau_k^2|\mathbf{y}, \mathbf{h}, \phi, \mu) \sim IG(\frac{N + \sigma_r}{2}, \frac{S_\sigma + \sum_{i=1}^{N} r_i^2}{2}) \tag{3.6}$$

where IG stands for an inverse gamma distribution, with shape and scale parameters, σ_r and S_σ. N is the number of time points in regime k, and r_i is the residual in regime k calculated from the volatility equation, that is, $r_i = h_{i+1} - \mu - \phi(h_i - \mu)$.

Inference for $\boldsymbol{\phi}$

Let $\phi_k = 2\phi^* - 1$, where ϕ^* has a Beta distribution with parameters (a, b). Then the prior distribution for ϕ_k is

$$f(\phi_k) \propto \{\frac{1 + \phi_k}{2}\}^{a-1} \{\frac{1 - \phi_k}{2}\}^{b-1} \tag{3.7}$$

The posterior distribution of ϕ_k is proportional to $f(\phi_k) f(\mathbf{h}|\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\tau})$ where $f(\mathbf{h}|\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\tau})$ is the product of normal distributions. A Metropolis Hasting sampling algorithm is used to sample from this full conditional distribution. For a discussion

Algorithm 1 Slice sampler to update $\boldsymbol{\mu}$

1: Assume the current value for μ_i is x. Sample $u \sim Unif(0, f(x))$. In practice the uniform draw is done in a log scale.
2: Solve the equation $f(x) = u$ which results in a quadratic equation; then take the smaller solution as a lower bound (LB_i), and the larger one as an upper bound (UB_i).
3: Sample $\mu_i \sim Unif(\max(LB_i, \mu_{i-1}), \min(UB_i, \mu_{i+1}))$.

of the Metropolis -Hastings algorithm, see [32].

Inference for $\boldsymbol{\mu}$

In a non-regime switching stochastic volatility model, a diffuse prior is assumed and the full conditional distribution for μ is [32]

$$f(\mu|\mathbf{h}, \phi, \tau) \sim N(\hat{\mu}, \sigma_\mu^2) \tag{3.8}$$

where

$$\hat{\mu} = \sigma_\mu^2\{\frac{1-\phi^2}{\tau^2}h_1 + \frac{1-\phi}{\tau^2}\sum_{i=1}^{N-1}(h_i - \phi h_{i-1})\} \tag{3.9}$$

$$\sigma_\mu^2 = \tau^2\{(n-1)(1-\phi^2) + (1-\phi^2)\}^{-1} \tag{3.10}$$

However, in our regime switching framework, we have an order for $\boldsymbol{\mu}$. Therefore an independent Gibbs sampler for each μ_k from its full conditional distribution is not possible. We resort to a slice sampler where we can sample from the full conditional distribution of μ_k with an upper bound and/or lower bound imposed. Note that the density of a normal distribution can be written as

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(x-\mu)^2}{2\sigma^2} \tag{3.11}$$

and the slice sampling algorithm that we implement is presented in Algorithm 1.

3.4.2 Inference for the Hidden Volatilities

Sample h one at a time

Conditional on $\mathbf{D}$ and $\boldsymbol{\theta}$, the full conditional distribution of $\mathbf{h}$ is the same as in a non regime switching stochastic volatility model. Therefore for simplicity, we omit $\mathbf{D}$ in this subsection. Because of the high dimensionality and autocorrelation,

sampling from the distribution $(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$ is more complicated than sampling from the distribution of $\boldsymbol{\theta}$. In the literature, Jacquier [29] samples each element of $\mathbf{h}$ from its full conditional distribution

$$p(h_t|\mathbf{h}_{-t}, \mathbf{y}, \boldsymbol{\theta}) = p(h_t|h_{t-1}, h_{t+1}, \boldsymbol{\theta}, \mathbf{y}) \tag{3.12}$$

where $\mathbf{h}_{-t}$ is the series $\mathbf{h}$ without h_t.

There are a number of competing strategies for sampling $\mathbf{h}$, see for example, Jacquier, Polson and Rossi [29], Shephard and Kim [32], Geweke [19] and Shephard and Pitt [45]. Jacquier, Polson and Rossi [29] propose to use an accept/reject Metropolis algorithm while in most other cases, an acceptance-rejection sampling method is enough.

Sample h simultaneously

Since $\{h_t\}$ are highly correlated, the initial updating schemes in the literature ware not efficient in order to overcome the problem of autocorrelation. Kim (1998)[32] suggests a transformation of $\mathbf{y}$ to $\mathbf{y}^*$ and using an offset mixture representation to approximate the distribution of the error term for y_t^*. Such an representation allows for filtering based methods which can generate $\mathbf{h}$ simultaneously from a multivariate distribution. To be explicit, let

$$y_t^* = h_t + z_t \tag{3.13}$$

where $y_t^* = log(y_t^2) + c$ and z_t follows a log χ^2 distribution.

$$f(z_t) = \Sigma_{i=1}^{L} q_i f_N(z_i|m_i - 1.2704, \nu_i^2). \tag{3.14}$$

is a mixture of L normal densities f_N. The means and variances of the normal densities are denoted by $m_i - 1.2704$ and ν_i^2. The component probabilities are given by q_i. The constants L and m_i, q_i, ν_i^2 are selected to approximate the exact distribution of z_t. According to Kim [32], $L = 7$ is the most efficient choice, and the parameters used for this mixture of normal densities are given in table 3.1.

This approximating mixture density can be written in terms of a component

ω	$Pr(\omega = i)$	m_i	ν_i^2
1	0.00730	-10.129999	5.79596
2	0.10556	-3.97281	2.61369
3	0.00002	-8.56686	5.17950
4	0.04395	2.77786	0.16735
5	0.34001	0.61942	0.64009
6	0.24566	1.79518	0.34023
7	0.25750	-1.08819	1.26261

Table 3.1. Mixing distribution to approximate $\log \chi^2$

indicator variable s_t such that

$$z_t|s_t = i \sim N(z_i|m_i - 1.2704, \nu_i^2), Pr(s_t = i) = q_i. \tag{3.15}$$

The previous parameter space is then augmented with $\mathbf{s}$, and $\mathbf{s}$ can be sampled independently from its full conditional distribution given below,

$$Pr(s_t = i|y_t^*, h_t) \propto q_i f_N(y_t^*|h_t + m_i - 1.2704, \nu_i^2). \tag{3.16}$$

After the transformation to $\mathbf{y}^*$, the model can be placed into a state space structure where a Kalman filter can be used to sample from the multivariate Gaussian distribution $\mathbf{h}|\mathbf{y}^*, \mathbf{s}, \boldsymbol{\theta}$ simultaneously. A description of the sampling algorithm proposed by Jong and Shephard [15] is given in Appendix A.

3.4.3 Inference for the Latent Regime Process

In our model, each realization of $\mathbf{D}$ is characterized by a series of jump chains and waiting times between jumps, which we call intervals. When the length of each interval follows an exponential distribution, $\mathbf{D}$ is Markovian and it is equivalent to a continuous-time Markov Switching Model. This special case and its estimation algorithms have been discussed in [36]. We generalize the waiting time distribution to be gamma, which gives rise to a non-Markovian property of $\mathbf{D}$. To our knowledge, it is the first time a non-Markovian regime structure is allowed in regime switching stochastic volatility models, and this generalization should be able to be applied in any other regime switching models. We now describe the estimation

Algorithm 2 Reversible jump MCMC to update $\mathbf{D}$

1: **PROPOSE** a realization of $\mathbf{D}_{new}$ based on current $\mathbf{D}_{old}$ using one of the following
   Birth Death Sampler
   Refinement Sampler
   Independent Sampler
2: **Calculate** the acceptance probability ψ
3: **ACCEPT** the proposed realization with probability ψ

procedure for $\mathbf{D}$ in detail.

Because the number of parameters to characterize a realization of $\mathbf{D}$ changes, we use a reversible jump MCMC algorithm to update $\mathbf{D}$ in each sweep. There are two steps in a reversible jump MCMC as listed in Algorithm 2:

Following the definition in [36], we describe the three samplers and their proposal probability $q(\mathbf{D}_{new}|\mathbf{D}_{old})$ below. Define $\mathbf{D}_{old}$ as the current realization of the regime process and $\mathbf{D}_{new}$ as the proposed realization. Assume there are n intervals in one realization, then the jump times are labeled as $i_1, ...i_n$ and the waiting times are labeled as $t_1, ...t_n$. $T_k = \sum_{j=1}^{k} t_j, k = 1, ..., n$ is the cumulative waiting time until the kth jump. The first interval $[0, T_1]$ and the last interval $[T_n, N]$ are called external intervals and others are called internal intervals.

Birth-Death Sampler

A birth-death sampler proposes a new realization of $\mathbf{D}_{new}$ by creating a new interval in the current realization of $\mathbf{D}_{old}$ or by removing an exiting interval $\mathbf{D}_{old}$. The details are listed in Algorithm 3.

Refinement Sampler

A refinement sampler proposes to adjust the location of one of the jump times in current realization $\mathbf{D}_{old}$. The details are listed in Algorithm 4.

Independent Sampler

An independent sampler generates a completely new realization of $\mathbf{D}_{new}$ randomly. Such a proposal has a very low acceptance ratio in practice.

Acceptance Probability

The acceptance probability is

$$\begin{aligned} \psi &= \frac{p(\mathbf{D}_{new}|-)q(\mathbf{D}_{old}|\mathbf{D}_{new})}{p(\mathbf{D}_{old}|-)q(\mathbf{D}_{new}|\mathbf{D}_{old})}, \\ p(\mathbf{D}_{new}|-) &\propto f(\mathbf{h}|\mathbf{D}_{new}, \boldsymbol{\Theta}) f(\mathbf{D}_{new}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{P}), \end{aligned} \tag{3.17}$$

Algorithm 3 Birth-Death Sampler

1: Select a birth with probability 1/2; otherwise select a death
2: Select an interval to be modified by randomly picking a time T' uniformly form $\mathbf{D}_{old}$. The interval $[T_i, T_{i+1}]$ containing T' will be modified
3: If a birth is selected, with equal probabilities create an interval using:
    * Middle Birth: Randomly pick a time T'' from $[T_i, T_{i+1}]$; the sub-interval $[min(T', T''), max(T', T'')]$ will be the new interval
    * Left Birth, the interval $[T_i, T']$ will be the new interval
    * Right Birth, the interval $[T', T_{i+1}]$ will be the new interval
    * Choose a new state excluding the state of the selected interval and assign the new state to the new interval. If the new state is the same as the state of the neighboring interval of the new interval, such a birth sampler is equivalent to a refinement sampler.
4: If a death is selected:
    * If the first interval is selected, absorb it into the second interval
    * If the last interval is selected, absorb it into the next to last interval
    * Otherwise, with equal probability absorb it into its immediate proceeding or following interval

Algorithm 4 Refinement Sampler

1: Randomly pick a point T' uniformly in $\mathbf{D}_{old}$
2: Change one of the jump times of an interval closest to T' as follows
    If T' falls on the first interval, then $T_1 = T'$
    If T' falls on the last interval, then $T_n = T'$
    If T' falls on the interval $[T_i, T_{i+1}]$, let $T_i = T'$ with probability 0.5; otherwise $T_{i+1} = T'$

where $p(\mathbf{D}_{new}|-)$ is the full conditional distribution of $\mathbf{D}_{new}$. $q(\mathbf{D}_{new}|\mathbf{D}_{old})$ is the proposal probability which depends on the type of sampler we use to propose the new realization $\mathbf{D}_{new}$ [36].

3.4.4 Inference for the Transition Matrix

Given that a jump occurs, the probability that $\mathbf{D}$ changes from state i to state j is p_{ij}, which is given by the jump transition matrix $\mathbf{P}$,

$$\begin{pmatrix} p_{11} & p_{12} & \dots & p_{1K} \\ p_{21} & p_{22} & \dots & p_{2K} \\ \dots & \dots & \dots & \dots \\ p_{K1} & p_{K2} & \dots & p_{KK} \end{pmatrix}$$

where $p_{ij} = Pr(D_{t+1} = j | D_t = i)$ with $\sum_{j=1}^{K} p_{ij} = 1$.

Each row of the matrix follows a Dirichlet prior distribution and the posterior distribution is multinomial. In practice, a gamma distribution is used to approximate the posterior multinomial distribution [30].

3.4.5 Inference for the Waiting Time Parameters

The intervals of **D** which belong to the i^{th} regime follow a gamma $G(\alpha_i, \beta_i)$ distribution. Gamma prior distributions are imposed so that $\alpha_i \sim G(a_1, b_1)$ and $\beta_i \sim G(a_2, b_2)$.

The posterior density for β_i is a conjugate gamma distribution:

$$\begin{aligned} f(\beta_i) &\propto p(\beta_i|-)G(a_2, b_2), \\ &\sim G(N_i\alpha_i + a_2 - 1, b_2 \sum_{j \in R_i} t_j) \end{aligned} \tag{3.18}$$

where R_i means in the i^{th} regime. N_i is the number of intervals in regime i and $p(\beta_i|-)$ is the full conditional distribution of β_i. It is noteworthy that when the last interval falls in the i^{th} regime, the posterior probability of β_i is a little bit different because we use a cumulative density for the last interval in the full conditional distribution calculation.

The sampling for the shape parameter α_i is more complicated due to the non-conjugate posterior probability distribution. We resort to a Metropolis-Hastings sampler. The posterior density of α_i is

$$\begin{aligned} f(\alpha_i) &\propto p(\alpha_i|-)G(a_1, b_1), \\ &\propto (\beta_i^{N_i} \prod_{j \in R_i} t_j)^{\alpha_i} \alpha_i^{a_1} exp(-\alpha_i b_1)/\Gamma(\alpha_i) \end{aligned} \tag{3.19}$$

For any interval j in regime i, that is, $j \in R_i$, we have $t_j \sim G(\alpha_i, \beta_i)$. Given an updated value β_i, according to the moments of a gamma distribution, the method of moment estimate for α_i is $m_i = \beta_i \sum_{j \in R_i} t_j / N_i$. We propose a new value $\alpha_i' \sim N(m_i, c)$, where c is a constant tuning parameter. Such a proposal is

Algorithm 5 Univariate Regime Switching Stochastic Volatility Model Initialization

1: Start with a basic stochastic volatility model where there is no regime switching to obtain a reasonably good sample of $\mathbf{h}$ and $\mathbf{s}$.
2: Set the initial values of $\boldsymbol{\mu}$ by dividing up the posterior distribution of $\boldsymbol{\mu}$ from the basic SV model. For example, when $K = 2$, set $\mu_1 = \overline{\mathbf{h}} - c*\sigma(\mathbf{h})$ and $\mu_2 = \overline{\mathbf{h}} + c*\sigma(\mathbf{h})$ where $\overline{\mathbf{h}}$ and $\sigma(\mathbf{h})$ are the mean and standard deviation of $\mathbf{h}$ obtained in the first step. c is a constant.
3: Fix $\mathbf{h}$ and $\boldsymbol{\mu}$ and update $\mathbf{D}$ and other parameters for certain number of iterations.
4: Start running the full algorithm with these initial values.

accepted with an acceptance ratio

$$\psi = \frac{f(\alpha_i')N(\alpha_i^c|m_i, c)}{f(\alpha_i^c)N(\alpha_i'|m_i, c)} \tag{3.20}$$

Once again, caution should be taken for the regime which contains the last interval.

3.4.6 Initialization

There are two aspects of the sampling strategy that are worth noting. First, because the last interval of $\mathbf{D}$ is censored, when calculating the likelihood for a realization of $f(\mathbf{D}'|\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{P})$, the probability of this interval should be the cumulative density $p(X \geq x)$, where x is the length of the last interval.

The second one is of a computation concern. Good starting values of the parameters can greatly expedite the convergence rate. Therefore careful considerations should be taken to find a reasonable initialization strategy for our model. In practice, we propose a initialization strategy which is described in Algorithm 5.

3.4.7 Parallel Tempering to Improve Mixing

The target joint distribution of parameters, $f(\mathbf{h}, \mathbf{D}, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{P}|\mathbf{y})$ may have multiple local maxima which trap MCMC updates so that we can not traverse the whole parameter space freely. To address this problem, we adopt a parallel tempering algorithm to improve mixing [20]. The basic idea of this algorithm is to allow for

M parallel chains to run our proposed MCMC updates at the same time. These chains will then swap their information by exchanging updated parameters. Each chain has a pre-specified "temperature" which determines the smoothness of the joint distribution of parameters. To be more specific, the joint distribution for the K^{th} chain is tempered to be $f(\mathbf{\Theta}|\mathbf{y})^{1/\lambda_K}$, where λ_K is the temperature for the K^{th} chain. In our exercise we number the chains by $K, K = 1, ..., M$ and their associated temperatures have an increasing order, starting from $\lambda_1 = 1$, or there is no tempering for the first chain. The temperatures are fixed throughout the MCMC analysis. In each sweep of the MCMC algorithm we update each chain separately. Then two chains, K_1, K_2 are selected to swap all the parameters except the temperatures. A swap will be accepted with an acceptance ratio

$$\alpha = \frac{f_{\lambda_{K1}}(\mathbf{\Theta}_{K2}|\mathbf{y}) f_{\lambda_{K2}}(\mathbf{\Theta}_{K1}|\mathbf{y})}{f_{\lambda_{K2}}(\mathbf{\Theta}_{K2}|\mathbf{y}) f_{\lambda_{K1}}(\mathbf{\Theta}_{K1}|\mathbf{y})} \tag{3.21}$$

3.5 Model Comparison

As mentioned before, our model is very flexible. The number of regimes is not restricted to be two. We can also choose to equate the persistence and variance parameters for the volatility process across different regimes, or allow them to be different. In this regards, good model comparison tools are necessary to choose from multiple model candidates. In this study, we utilize four model comparison criteria. These criteria may arrive at different conclusions because they are designed to answer different questions. For example, the Bayes factor based methods "Harmonic Mean" and "Filter" focus more on the predictability of prior information, while DIC addresses how well the posterior predicts future data [5].

3.5.1 Bayes Factor

The marginal likelihood of our model is denoted as $m(\mathbf{y})$ and

$$m(\mathbf{y}) = \int f(\mathbf{y}|\mathbf{\Theta}) f(\mathbf{\Theta}) d\mathbf{\Theta} \tag{3.22}$$

where $\mathbf{\Theta}$ includes all the parameters in the augmented parameters space from our model. The Bayes factor comparing two models, M_1 and M_2, is given by $\frac{m(\mathbf{y}|M_1)}{m(\mathbf{y}|M_2)}$.

A value greater than one gives preference to model M_1 over model M_2. The idea is that a model with a larger marginal likelihood fits the data better and is more likely to be the true model. In our study, instead of providing a Bayes factor for each pair of the models, we calculate the marginal likelihood for each model, and choose the one with the largest value. A Bayes factor is easily computed given the marginal likelihoods.

Because of the high dimensionality of our parameter space, computing the marginal likelihood by integration is not straightforward and we use two methods to approximate this integral.

Harmonic Mean

Using an importance sampling argument, Newton & Raftery [42] proposed a simulation based marginal likelihood calculation method by taking the harmonic mean of $f(\mathbf{y}|\mathbf{\Theta}^{(i)})$ where $\mathbf{\Theta}^{(i)}$ is the updated parameters from the i^{th} MCMC sweep. Assume M is the number of total MCMC iterations, and define the marginal likelihood estimator $\tilde{m}$ as:

$$\hat{m}(\mathbf{y}) = \{\frac{1}{M}\sum_{i=1}^{M}\frac{1}{f(\mathbf{y}|\mathbf{\Theta}^{(i)})}\}^{-1} \tag{3.23}$$

where

$$f(\mathbf{y}|\mathbf{\Theta}^{(i)}) = \prod_{t=1}^{N}\frac{1}{\sqrt{2\pi\exp(h_t^{(i)})}}\exp -\frac{y_t^2}{2\exp(h_t^{(i)})} \tag{3.24}$$

Filtering

Another popular method to calculate the marginal likelihood is based on an augmented auxiliary particle filter proposed in [13] and [45]. The objective of the filter is to calculate the marginal log-likelihood $\log f(\mathbf{y})$ successively by

$$\log f(\mathbf{y}) = \log(y_1) + \sum_{t=2}^{N}\log f(y_t|\mathbf{y}_{t-1}) \tag{3.25}$$

where $\mathbf{y}_{t-1} = (y_1, ..., y_{t-1})$. We describe the detailed procedure in Appendix B under a non regime switching framework.

To accommodate the regime switches in our model, the simulation based sequential analysis introduced in [11], which is an extension of the algorithm in the

appendix, is directly applicable.

One disadvantage of this method is that it is based on the Markovian assumption of the hidden regime process $\mathbf{D}$, which does not hold for the non Markovian regime switching case where the waiting time follows a gamma distribution. As a result, we cannot compare gamma interval models based on the filtering method.

3.5.2 DIC

We also employ a DIC model comparison method, which addresses how well the posterior from a particular model can predict future data, and has the advantage of being computationally simple, see [5]. DIC is defined as

$$DIC = \bar{D} + p_D \tag{3.26}$$

where the first term is the deviance, which measures the goodness of fit, and the second term penalizes complex models.

$$\bar{D} = E_{\mathbf{\Theta}|\mathbf{y}}[D(\mathbf{\Theta})] = E_{\mathbf{\Theta}|\mathbf{y}}[-2lnf(\mathbf{y}|\mathbf{\Theta})]. \tag{3.27}$$

$$p_D = \bar{D} - D(\bar{\mathbf{\Theta}}) \tag{3.28}$$

where $\bar{\mathbf{\Theta}}$ is the posterior mean of parameters. Therefore, p_D is the difference between the posterior mean deviance and the deviance at the posterior parameter means. A low DIC value indicates a better model fit.

3.5.3 Baby Reversal Jump algorithm

Reversible Jump MCMC is useful for exploring and comparing models with different parameter spaces [21][37][35]. We use it to calculate the posterior probability of different models. To be more specific, assume we have K models and a uniform prior is placed over the K model space. Let $\mathbf{\Theta}_k, \quad k = 1, ..., K$ denote the parameters under the k^{th} model. Assume the current model is k_1 with parameters $\mathbf{\Theta}_{k1}$. Begin by proposing a candidate model k_2 and update the parameters $\mathbf{\Theta}_{k2}$ for model k_2 based on our MCMC algorithms described in the previous sections. Such a move to model k_2 leads to a Metropolis-Hastings acceptance probability

which, after calculation, can be simplified as a ratio of likelihoods from the two competing models

$$\alpha = \min\{1, \frac{p(\mathbf{y}|\mathbf{\Theta}_{k2})}{p(\mathbf{y}|\mathbf{\Theta}_{k1})}\} \tag{3.29}$$

where we accept model k_2 with probability α. While iterating, we count the number of times of each model is selected, which is proportional to the posterior probability for each model, indicating the fitness of the model of the data set.

3.6 Simulation

We generate two sets of log return time series, one from a Markovian regime switching process and the other from a non-Markovian regime switching process. The insights from the simulation study are twofold: first, for both data sets, our proposed MCMC algorithm converges and provides accurate parameter estimations; and second, the model comparison tools are able to identify the true model out of various competing models.

3.6.1 Markovian Regime Process

We simulate $N = 5000$ data points according to a 2-regime continuous-time regime switching stochastic volatility model. For the generated data, regimes are distinguished only by different mean volatility level $\mu_1 \neq \mu_2$, while persistency and variance of volatilities are the same across different regimes ($\phi_1 = \phi_2$, $\tau_1 = \tau_2$). The waiting times follow an exponential distribution with parameter $\alpha = 1$, $\beta = 0.005$, that is, we generate a Markovian volatility regime process. For the volatility parameters, we follow the prior distributions specified in [32] and for the regime process parameters, we provide uninformative conjugate priors, which are summarized in Table 3.2.

The simulated log returns $\mathbf{y}$ and log volatilities $\mathbf{h}$ are presented in Figure 3.1. We start the model fitting with a non-regime switching model to obtain a reasonable initial value for $\mathbf{h}$, and using this initialization, we implement the full MCMC algorithm. We discard the first 100,000 burn-ins and for the following 300,000 iterations, we store every 20^{th} iteration.

The complete stored 15,000 iterations of the four parameters, $\mu_1, \mu_2, \phi, \tau^2$, are

graphed in Figure 3.3 and summarized in Table 3.2. It is clear from Figure 3.3 that the parameters have converged against iterations, the posterior density plot of each parameter is well shaped and the algorithm is efficient because the sample autocorrelations decay fast. The comparison between true parameters and their posterior mean and posterior standard deviations in Table 3.2 shows that our algorithm provides accurate parameter estimates. As for the hidden regime process, the first panel in Figure 3.2 presents true **D** and the second and third panels provide the posterior probabilities for each time point of being in regime 1 and regime 2 respectively. This figure indicates that the hidden regime process **D** has been successfully recovered.

We also compare several competing models using this simulated data. The models are summarized in Table 3.3. Model 1 is the true model; model 2 is a benchmark simple stochastic volatility model without regime switching. The remaining models, 3-5, are all 2-regime switching stochastic volatility model with exponential waiting times. These models differ in whether the volatility parameters ϕ and τ are the same or change across regimes. In model 6 we assume 3 regimes and in model 7 we assume two regimes but a gamma waiting time for the regimes. For both of the models we assume equal ϕ and τ across regimes. The model comparison results are shown in Table 3.4. As anticipated, the true model has the largest marginal likelihood, smallest DIC and a posterior model probability of essentially one. All the four criteria identify the true model as the best fitting model, even compared to other more highly parameterized models.

Parameter	True Value	Estimates	Prior
μ_1	-1.3	-1.29 (0.06)	uniform
μ_2	0.2	0.22(0.04)	uniform
$\phi_1 = \phi_2$	0.6	0.59(0.07)	$0.5 * (\phi + 1) \sim Beta(20, 1.5)$
$\tau_1^2 = \tau_2^2$	0.2	0.23(0.04)	$\tau^2 \sim IG(2.5, 0.025)$
$\alpha_1 = \alpha_2$	1	1	constant, exponential distribution
β_1	0.005	0.005(0.001)	$\beta_1 \sim Gamma(0.1, 0.01)$
β_2	0.005	0.006(0.001)	$\beta_2 \sim Gamma(0.1, 0.01)$

Table 3.2. True parameters, posterior parameter means and posterior standard deviations (in parenthesis) and prior distributions used for univariate stochastic volatility simulation

Model	Total Regimes	Waiting Time	$\mu_1 \neq \mu_2$?	$\phi_1 \neq \phi_2$?	$\sigma_1^2 \neq \sigma_2^2$?
1	**2**	**EXP**	**YES**	**NO**	**NO**
2	1	–	–	–	–
3	2	EXP	YES	YES	NO
4	2	EXP	YES	NO	YES
5	2	EXP	YES	YES	YES
6	3	EXP	YES	NO	NO
7	2	GAMMA	YES	NO	NO

Table 3.3. Different models compared for simulated Markovian univariate data. Note model 1 is the true model, or the model used to generate synthetic data.

Model	Harmonic Mean	Filter	DIC	Posterior Prob
True Model	**-5530**	**-5956**	**11533**	**1**
No regime	-5757	-6012	11778	0
μ, ϕ differ	-5540	-5957	11548	0
μ, σ^2 differ	-5543	-5959	11542	0
All differ	-5551	-5960	11570	0
Three regime, μ differ	-5537	-5960	11568	0
Gamma Interval	-5542	*	11555	0

Table 3.4. Model comparison results by different criteria for simulated Markovian univariate volatility data

3.6.2 Non-Markovian Regime Process

Generalizing **D** to be a non-Markovian regime process, we simulate data where the waiting times follow a gamma distribution. To be specific, we simulate a time series with the same parameters as in Section 3.6.1 except that the waiting times follow a gamma $G(20, 0.1)$ distribution, which makes the hidden regime process **D** non-Markovian.

The generated series **y** and the volatility **h** are shown in Figure 3.4. This figure suggests there are fairly regular switches in the volatility regimes. We examine the performance of the proposed model using both Markovian and non-Markovian regime switching models. The model comparison results are shown in Table 3.5. We leave $*$ for the "Filter" column because the filter method is not applicable to non-Markovian regime processes. Again as anticipated, the non-Markovian model has a larger Harmonic Mean, smaller DIC value and a posterior probability

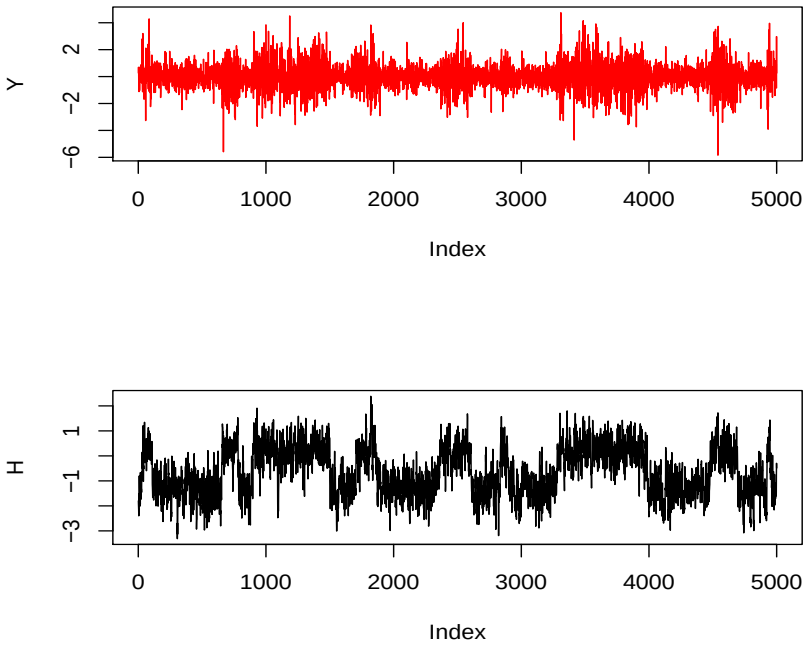

Figure 3.1. Simulated Markovian regime switching volatility data: upper plot: log return series; lower plot: log volatilities series

of essentially one for the correct model, all of which give preference to the non-Markovian model, confirming the ability of our model comparison tools to correctly identify the true model.

Model	Harmonic Mean	Filter	DIC	Posterior Prob
Markovian	-4856	*	10168	0
Non-Markovian	**-4847**	*	**10162**	**1**

Table 3.5. Model comparison between Markovian/non-Markovian regime switching stochastic volatility models for simulated non-Markovian data

3.7 Application to Real Data

To investigate the appropriateness of the proposed non-Markovian regime switching stochastic volatility model in real practice, we applied it to two types of real

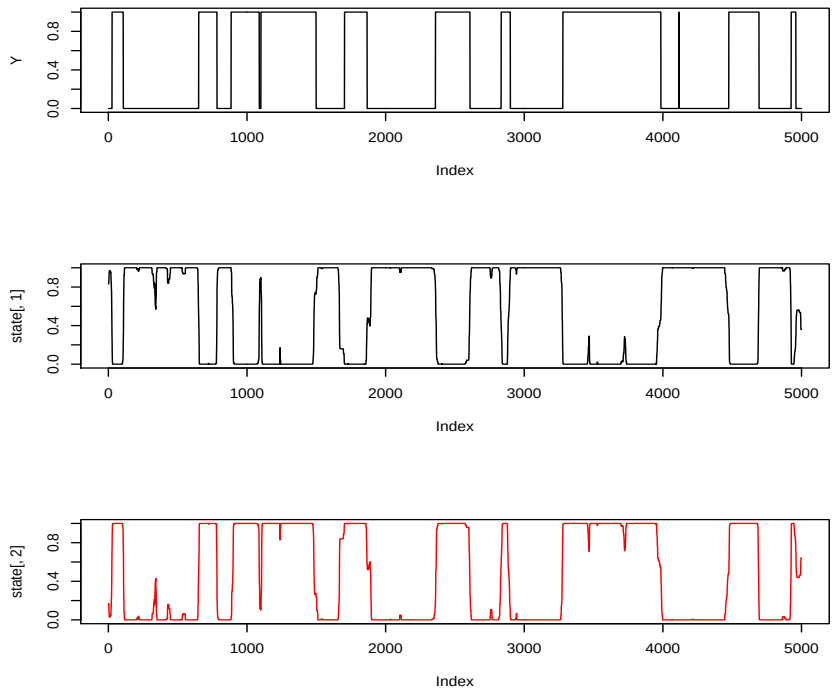

Figure 3.2. Probability of regimes: the first plot shows the true volatility regimes; the second shows the posterior probabilities of being in regime 1 and the third shows the posterior probabilities of being in regime 2

data: (I) data which may have regular switches in the volatility structures; and (II) data which do not have evident regular volatility switches. For a comparison, we also applied the Markovian regime switching model to both data sets. Model comparison results indicate that the non-Markovian model provides a better fit for type I data and Markovian model matches type II data.

3.7.1 Natural Gas Commodity

First we applied our models to daily Natural Gas future data which is a typical example of type I data. It ranges from 09/11/1997 to 03/16/2009, with $N = 3000$. In both cases, we assume two regimes ($K = 2$) and we allow $\mu_1 \neq \mu_2$, $\phi_1 \neq \phi_2$ and $\tau_1^2 \neq \tau_2^2$.

The posterior probability estimates for each regime for both the Markovian

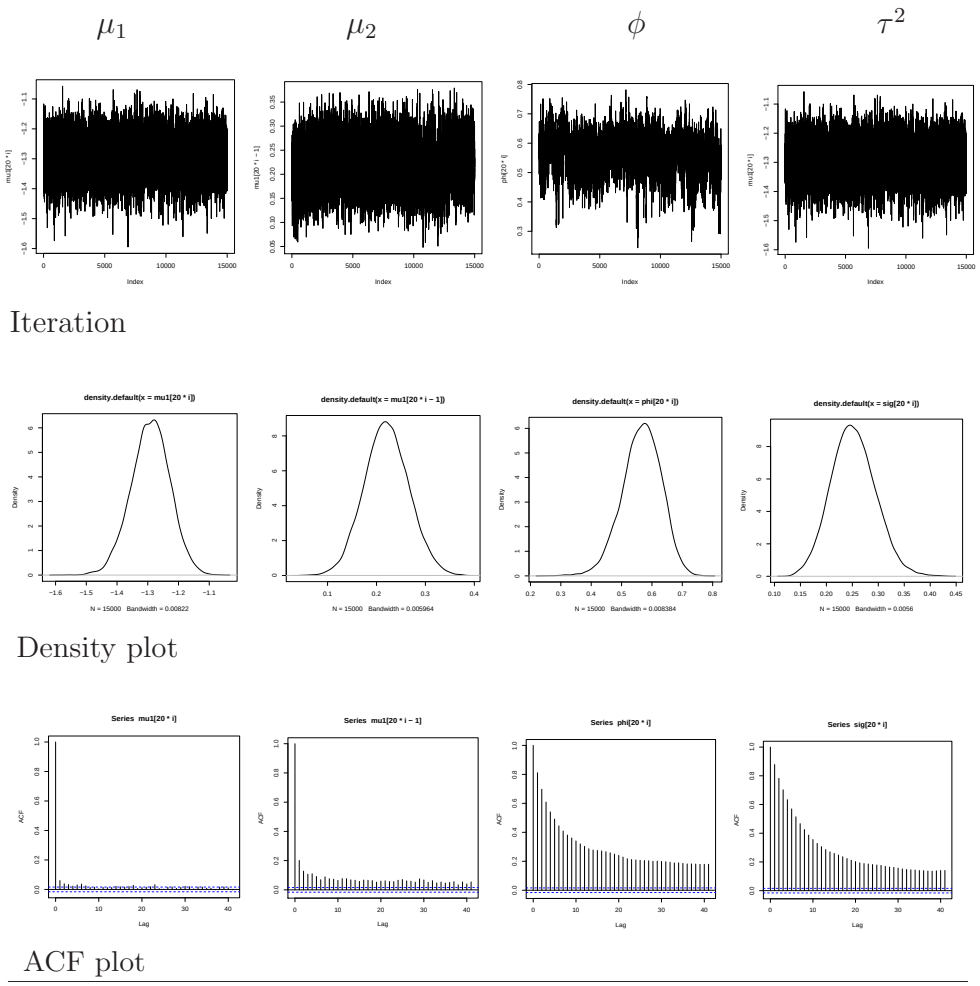

Figure 3.3. MCMC sampling results for simulation data: (row 1) trace plots; (row 2) histograms of posterior marginal distribution; (row 3) corresponding Autocorrelation plot (ACF) for each parameter trace plot.

and the non-Markovian model are displayed side by side in Figure 3.5. The plot of log returns suggests that this price process experiences periodic increases in the volatility, but neither the high nor the low volatility regimes last for long. The non-Markovian model captures this feature in that the posterior regime process switches states frequently while the Markovian model fails to find the finer structures and tends to identify longer intervals. All model comparison results (Table 3.6) also prefer a non-Markovian regime switching model to the regular Markovian model.

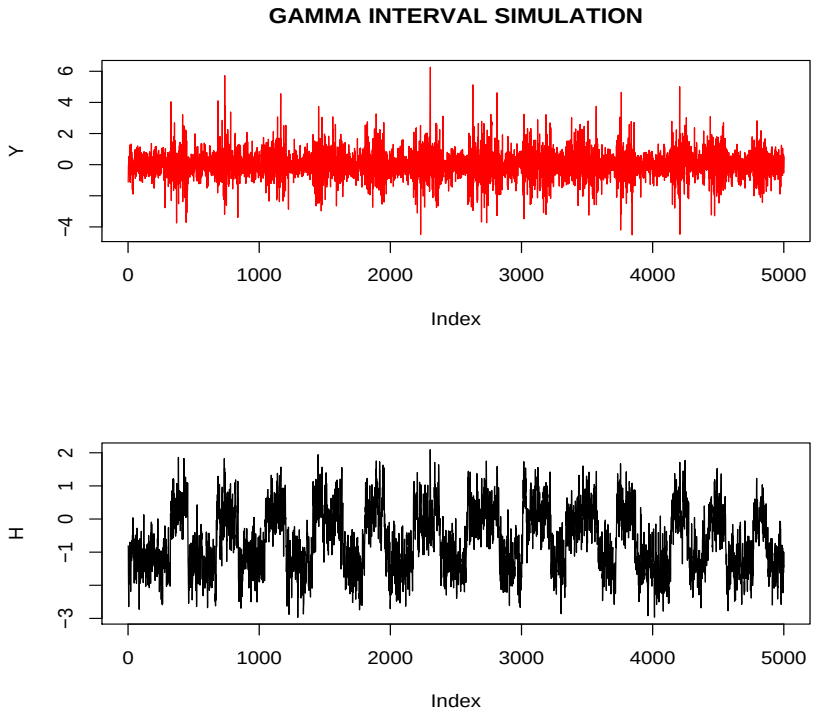

Figure 3.4. Simulated non-Markovian regime switching volatility data. upper plot: log return series; lower plot: log volatilities series

Model	Harmonic Mean	Filter	DIC	Posterior Prob
Markovian	-2764	*	5894	0
Non-Markovian	**-2732**	*	**5850**	**1**

Table 3.6. Model comparison results for natural gas commodity data: comparing Markovian/non-Markovian regime processes by different criteria

3.7.2 JPMorgan Stock Price

Although the non-Markovian switching model is superior for type I data which have regular structural switches in the volatilities, our empirical experiences tell that there are some cases where a Markovian model is sufficient. We therefore work with typical equity time series data: daily JPMorgan stock prices for the period 1996/08/29 to 2008/07/31 with $N = 3000$. The model comparison results are shown in Table 3.7 and for this data, the Markovian regime switching model is clearly preferred to the non-Markovian model.

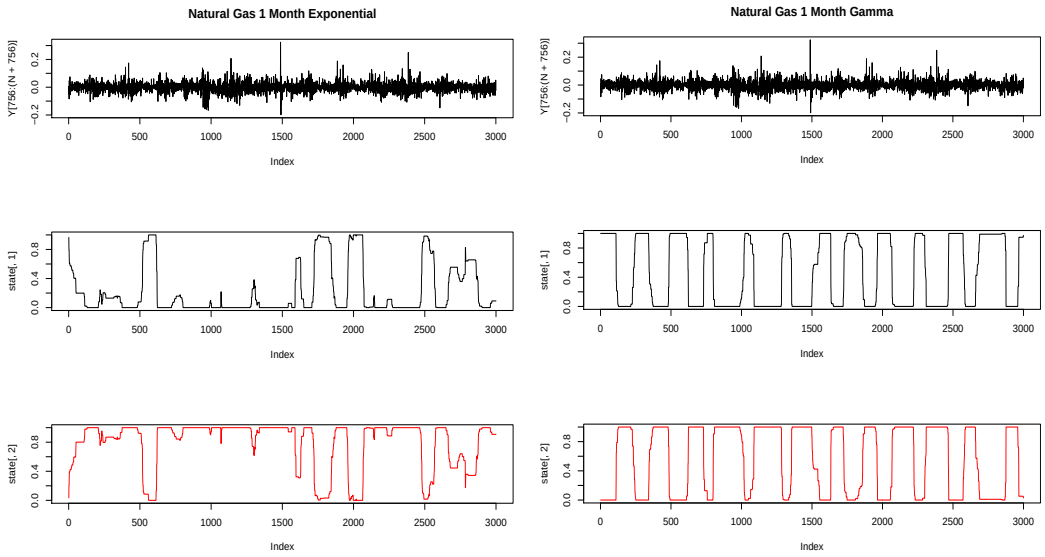

Figure 3.5. Probability of regimes for natural gas commodity data. The three plots in each panel are: log return data; posterior regime probabilities for regime 1; posterior regime probabilities for regime 2.

Model	Harmonic Mean	Filter	DIC	Posterior Prob
Markovian	**-5923**	*	**12107**	**1**
Non-Markovian	-5932	*	12167	0

Table 3.7. Model comparison results for JP Morgan stock price data: comparing Markovian/non-Markovian regime processes by different criteria

3.8 Conclusion for Univariate Regime Switching Stochastic Volatility Model

In this section, we proposed a continuous-time regime switching stochastic volatility model where the regime switching process **D** can be non-Markovian. The model is complex in that there are three layers of time series, two out of which are latent. We have the observable log return series **y**, unobservable log volatility series **h** and hidden regime process **D**. In order to efficiently estimate this model, we adopt a MCMC sampler where a Kalman filter based algorithm is used to sample **h** simultaneously and a reversible jump MCMC algorithm is used to sample **D**.

Simulation study shows that this algorithm is able to provide accurate parameter estimates.

Our model is flexible in that the number of regimes can take any value, the volatility parameters ϕ and τ^2 can be the same or different across regimes, and the regime switching process $\mathbf{D}$ can be Markovian or non-Markovian. The diversity of models provides us a platform to search for the best fitting model for a data set and to explore interesting econometric phenomena. We also provide several useful model comparison tools, which are shown to be effective in identifying the true model from various competing models using simulated data.

The most important feature of this model is that, to our knowledge, it is the first time that a regime switching process $\mathbf{D}$ is allowed to be non-Markovian in a continuous-time framework. To explore the properties of this extension, we work with both simulation and empirical data and find out that the proposed non-Markovian regime switching model is a better fit for data which possibly have periodic, or seasonal components, while a Markovian specification for $\mathbf{D}$ is sufficient for other types of data.

Chapter 4

Review of Multivariate Covariance Model

It is well documented that financial assets are correlated and do not evolve independently. For the last two decades, the study of volatility processes for multivariate time series has received considerable attention, spurred by the fundamental problems of portfolio allocation, risk management, and Value-at-Risk calculations.

As with the univariate volatility case, there are also two streams of research in the multivariate world: an extension of the GARCH framework and an extension of the Stochastic Volatility framework.

Yu and Meyer [51] summarized several empirical facts about returns of multiple financial assets:

- Returns are cross dependent
- Volatilities are cross dependent
- Volatility spills over from one market to another market
- There often exists a lower dimensional factor structure that can explain most of the correlation
- Correlations are changing over time

The following models in the multivariate covariance literature have tried to address these phenomena.

4.1 Multivariate GARCH Models and its Extensions

Early works focused mainly on multivariate GARCH models. A comprehensive survey of these works can be found in Laurent, Bauwens and Rombouts [34].

Assume there are p assets. The basic model is:

$$Y_t = H_t^{1/2} U_t \tag{4.1}$$

where Y_t is a vector of p asset returns and U_t is a random variable drawn from an i.i.d $N(0, I_p)$. The time varying covariance matrix H_t can be decomposed into

$$H_t = S_t \Gamma_t S_t \tag{4.2}$$

where S_t is a diagonal matrix of assets' time varying standard deviations and Γ_t is a time varying correlation matrix.

The log likelihood for this model can be written as

$$\begin{aligned} L &= -\frac{1}{2}\sum_{t=1}^{T}(p \cdot \log(2\pi) + \log(|H_t|) + Y_t^{'} H_t^{-1} Y_t), \\ &= -\frac{1}{2}\sum_{t=1}^{T}(p \cdot \log(2\pi) + \log(|S_t \Gamma_t S_t|) + Y_t^{'} S_t^{-1} \Gamma_t S_t^{-1} Y_t), \\ &= -\frac{1}{2}\sum_{t=1}^{T}(p \cdot \log(2\pi) + 2\log(|S_t|) + \log(|\Gamma_t|) + \widetilde{U}_t^{'} \Gamma_t \widetilde{U}_t). \end{aligned} \tag{4.3}$$

where $\widetilde{U}_t = S_t^{-1} Y_t = (\widetilde{u}_{1.t}, ..., \widetilde{u}_{K,t})^{'}$ is a zero mean process with covariance matrix Γ_t. The likelihood can be separated to two parts: a part for the standard deviations L_1, and conditional on the standard deviations, a part for correlations L_2, where

$$L_1 \propto 2\log(|S_t|) \tag{4.4}$$

$$L_2 \propto \log(|\Gamma_t|) + \widetilde{U}_t^{'} \Gamma_t \widetilde{U}_t. \tag{4.5}$$

Based on this separation, Bollerslev [8] proposed a two-step estimation method. In the first step, a set of independent univariate GARCH models are used to esti-

mate the volatilities for each asset. Then conditional on the extracted volatilities in step 1, the correlation matrix Γ_t can be estimated. There are three primary challenges in estimating correlation matrices: the large number of parameters for high dimensional correlations, the restriction that the correlation matrix is positive semi-definiteness (PSD) and the fact that correlations can be time varying. Many methodologies have been presented to address these challenges. In the following, we review the ones that are related to our research.

4.1.1 Constant Conditional Correlation GARCH Model

CCC-GARCH model was proposed by Bollerslev [8]. In this model,

$$S_t^2 = \text{diag}\{w_i\} + \text{diag}\{\kappa_i\} \circ Y_{t-1}Y_{t-1} + \text{diag}\{\lambda_i\} \circ S_{t-1}^2 \tag{4.6}$$

where S_t is the $p \times p$ diagonal matrix of assets' standard deviations at time t and each diagonal element in S_t follows a GARCH(1,1) pattern. The operator $\circ$ is the Hadamard product of two identically sized matrices, which is simply calculated by element wise multiplication. The correlation matrix Γ is symmetric positive definite matrix and assumed to be constant over time.

4.1.2 Dynamic Conditional Correlation GARCH Model

Robert Eagle presented a DCC-GARCH model in 2002 [17]. In his model, besides the volatilities, the correlation matrix also follows a GARCH type autoregressive relationship, or

$$\begin{aligned} \varepsilon_t &= S_t^{-1}Y_t, \\ Q_t &= C \circ (\mathbf{11'} - \mathbf{A} - \mathbf{B}) + \mathbf{A} \circ \varepsilon_{\mathbf{t-1}}\varepsilon'_{\mathbf{t-1}} + \mathbf{B} \circ \mathbf{Q_{t-1}}, \\ \Gamma_t &= \text{diag}\{Q_t^{-1}\}Q_t\text{diag}\{Q_t^{-1}\}. \end{aligned} \tag{4.7}$$

where the correlation matrix Γ_t is represented by a function Q. The matrix C is the unconditional correlation matrix of ε, the matrices A, B are constant parameter matrices and $\mathbf{1}$ is a vector of ones. It has been shown that if $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{11'} - \mathbf{A} - \mathbf{B}$ are positive semi-definite, then Q will be positive semi-definite.

4.1.3 Regime Switching Dynamic Correlation GARCH

The Regime Switching Dynamic Correlation GARCH model proposed by Pelletier [44] is a mid point between the CCC and the DCC models. This model incorporates a hidden regime switching process where the correlation matrix is assumed to be constant within a regime. By this definition, the correlation matrix is still time varying, but does not change at every time point as in the DCC model.

There are two versions of the regime switching DCC-GARCH model:

(1)The unrestricted version:

$$\Gamma_t = \sum_{k=1}^{K} I(D_t = k)\Gamma_k \tag{4.8}$$

where D_t, which can take K possible values, indicates the unobserved regime at time t and Γ_k is the correlation in regime k. A Choleski decomposition of the correlation matrix ensures the resulting matrix is PSD, and EM algorithm makes the high dimensional parameter estimations feasible.

(2)The restricted version:

$$\Gamma_t = \Gamma\lambda(D_t) + I_p(1 - \lambda(D_t)) \tag{4.9}$$

where Γ is a fixed correlation matrix, I_K is a $p \times p$ identity matrix, and $\lambda(D_t) \in [0, 1]$ is a univariate random process governed by an unobserved Markov chain D_t. Restrictions are needed for this model to make parameters identifiable. For example in this model, they require that $\lambda(1) = 1, \lambda(1) > \lambda(2) > ... > \lambda(K)$. One weakness with this model is that it does not allow the correlation matrix to change sign, which might happen in periods of market distress. This restriction, imposed through the assumption that $\lambda(D_t) \in [0, 1]$, is due to that fact that the authors do not have a result for a lower bound on $\lambda(D_t)$ that would guarantee that Γ is positive definite.

For the standard deviations, to ensure linearity, an ARMACH model is applied. In the ARMACH model, the conditional standard deviations follow the relationship below:

$$s_t = \omega + \sum_{i=1}^{q} \widetilde{\alpha}_i |q_{t-i}| + \sum_{i=1}^{p} \beta_j s_{t-j} \tag{4.10}$$

Because of linearity in this model, multi-step forecasting is available.

This modeling framework allowed the authors to address the interesting question of whether models with regime switching in the variance and models with regime switching in the correlation can be distinguished. They also confirmed that it is possible to differentiate and they proposed a model where both the variances and the correlations are driven by a switching regime process, but they did not implement this model in practice.

4.1.4 Markov Switching Dynamic Conditional Correlation GARCH Model

Markov Switching Dynamic Conditional Correlation GARCH model (Billio & Caporin [6]) extends the DCC-GARCH model by allowing both the unconditional correlation matrix C and parameter matrices $\mathbf{A}$, $\mathbf{B}$ in the correlation autoregressive equation to follow a regime switching process. Sharing the notations from previous sections, we have:

$$Q_t = C(D_t) \circ (\mathbf{11'} - \mathbf{A(D_t)} - \mathbf{B(D_t)}) + \mathbf{A(D_t)} \circ \varepsilon_{\mathbf{t-1}}\varepsilon'_{\mathbf{t-1}} + \mathbf{B(D_t)} \circ \mathbf{Q_{t-1}} \quad (4.11)$$

As in Pelletier's [44] model, this model imposes a regime switching structure only on the correlation matrix, leaving the standard deviation process to follow a standard non-regime switching model. The authors conclude that a full Markov switching model on both variance and correlation part will become very unstable due to large number of switching parameters.

4.2 Multivariate Stochastic Volatility (MSV) Model and its Extensions

The multivariate stochastic volatility model is a counterpart to the multivariate GARCH model. Because the likelihood functions of stochastic volatility models are not in a closed form, the parameter estimation is more difficult. Compared to the multivariate GARCH models, multivariate stochastic volatility models are new and less developed, offering a wider area for exploration. A summary of the

multivariate stochastic volatility models can be found in Chib, Omori and Asai [14]. Following the notations in Equation 4.1 and Equation 4.2, we provide a brief overview of multivariate stochastic volatility models.

4.2.1 Basic Multivariate Stochastic Volatility Model

The volatilities in a multivariate stochastic volatility model are not deterministic and as a result they follow a stochastic process. Let $S_t = \text{diag}(\exp(h_{1t}/2), ..., \exp(h_{pt}/2))$ and $\mathbf{h}_t = (h_{1t}, ..., h_{pt})$. Then

$$\mathbf{h}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t \tag{4.12}$$

where $\boldsymbol{\mu} = (\mu_1, ..., \mu_p)$ and $\boldsymbol{\Phi} = \text{diag}(\phi_1, ..., \phi_p)$ and $\boldsymbol{\eta}_t | \mathbf{h}_t \sim N(\mathbf{0}, \Sigma_{\eta\eta})$.

The correlation matrix Γ in the basic multivariate stochastic volatility model is assumed to be constant. However, analogous to the multivariate GARCH framework, the correlation matrix can be extended to be time varying.

4.2.2 Modeling by Reparameterization for the Bivariate Model

Yu and Meyer [51] proposed a bivariate stochastic volatility model with time varying correlations. If we denote the off-diagonal element in the 2×2 correlation matrix Γ_t as ρ_t, which is reparameterized by $\rho_t = \frac{\exp(q_t)-1}{\exp(q_t)+1}$ to ensure positive semi-definiteness (PSD), then we assume q_t follows a stochastic volatility type autoregressive relationship:

$$q_{t+1} = \psi_0 + \psi_1(q_t - \psi_0) + \sigma_\rho \nu_t, \quad \nu_t \sim N(0, 1) \tag{4.13}$$

While this is an interesting toy model, this reparameterization is difficult to generalize to higher dimensional data.

4.2.3 Dynamic Conditional Correlation - MSV model

The model proposed by Asai and McAleer [1] is a stochastic version of the DCC-GARCH model. Let $\mathbf{Q}_t$ be a sequence of positive definite matrices, and $\mathbf{Q}_t^*$ be a

diagonal matrix with the diagonals equal the diagonals of $\mathbf{Q}_t$. Then they define:

$$\begin{aligned}
\mathbf{\Gamma}_t &= \mathbf{Q}_t^{*-1/2}\mathbf{Q}_t\mathbf{Q}_t^{*-1/2}, \\
\mathbf{h}_{t+1} &= \widetilde{\boldsymbol{\mu}} + \mathbf{\Phi}\mathbf{h}_t + \boldsymbol{\eta_t}, \quad \mathbf{\Phi} = \mathrm{diag}(\phi_1, ..., \phi_p) \\
\boldsymbol{\eta_t} &\sim N_p(\mathbf{0}, \mathbf{\Sigma}_{\boldsymbol{\eta\eta}}), \\
\mathbf{Q}_{t+1} &= (1-\psi)\overline{\mathbf{Q}} + \psi\mathbf{Q}_t + \mathbf{\Xi}_t, \\
\mathbf{\Xi}_t &\sim W_p(\nu, \mathbf{\Lambda}).
\end{aligned} \tag{4.14}$$

where W_p is a p-dimensional Wishart distribution.

The DCC-MSV model differs from the DCC-GARCH model in two aspects. First, the volatility processes are stochastic, and possibly correlated if we allow the off-diagonal elements of the matrix $\mathbf{\Sigma}_{\boldsymbol{\eta\eta}}$ be nonzero. Second, the correlation matrices are allowed to change over time and are governed by a stochastic process, with disturbance $\mathbf{\Xi}_t$ following a Wishart distribution.

4.2.4 Factor MSV Model

Empirical results suggest that there often exists a lower dimensional factor structure that can explain most of the observed correlation in the financial markets. This approach forms the basis of the factor MSV model.

The basic factor model is represented as:

$$\begin{aligned}
Y_t &= \mathbf{B}\mathbf{f}_t + \mathbf{V}_t^{1/2}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N_p(\mathbf{0}, \mathbf{I}) \\
\mathbf{f}_t &= \mathbf{E}_t^{1/2}\boldsymbol{\gamma}_t, \quad \boldsymbol{\gamma}_t \sim N_q(\mathbf{0}, \mathbf{I}) \\
\mathbf{h}_{t+1} &= \boldsymbol{\mu} + \mathbf{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N_{p+q}(\mathbf{0}, \mathbf{I})
\end{aligned} \tag{4.15}$$

$$\mathbf{V}_t = \mathrm{diag}(\exp(h_{1t}), ..., \exp(h_{pt})) \tag{4.16}$$

$$\mathbf{E}_t = \mathrm{diag}(\exp(h_{p+1,t}), ..., \exp(h_{p+q,t})) \tag{4.17}$$

where $\mathbf{f}_t$ is a vector of lower dimensional (q dimension) common factors which can explain most of the correlation of Y_t. The matrix $\mathbf{B}_t$ is $p \times q$ loading matrix, which describes how much each financial asset is driven by a specific underlying common factor. This approach has the advantage that it tackles the high dimension problem

of multivariate models.

The covariance matrix for Y_t is $\mathbf{B}\mathbf{D}_t\mathbf{B} + \mathbf{V}_t$. Because $\mathbf{h}_t$, which constitutes the covariance component $\mathbf{E}_t$ and $\mathbf{V}_t$, is a $p+q$ vector that follows a stochastic autoregressive process, the covariance matrix is also time varying.

A recent paper by Lopesa and Carvalho [38] has extended the basic factor model along two directions as described below.

Time Varying Factor Loading B

The first extension allows the loading matrix $\mathbf{B}$ to be time varying. Stacking up the elements of the matrix $\mathbf{B}$, we can get a $d = pq - q(q-1)/2$ long vector $\boldsymbol{\beta}$. Assume that $\boldsymbol{\beta}$ follows a first order autoregressive process:

$$\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1}, \boldsymbol{\zeta}, \boldsymbol{\Theta}, \mathbf{W}_t \sim N(\boldsymbol{\zeta} + \boldsymbol{\Theta}\boldsymbol{\beta}_{t-1}; \mathbf{W}_t) \tag{4.18}$$

where the evolution of $\mathbf{W}_t$ is completely specified by a single, known discount factor $\delta \in (0, 1)$. Given this specification, the time varying property of the covariance matrix is not explained solely by $\mathbf{h}$ anymore.

Regime Switching Factors

The second extension (Lopesa and Carvalho)[38] allows the factors $\mathbf{h}' = (h_{p+1}, ..., h_{p+q})$ to follow a regime switching stochastic volatility type model. That is,

$$\mathbf{h}_{t+1} = \boldsymbol{\mu}_{D_t} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}_{D_t}) + \boldsymbol{\eta}_t \tag{4.19}$$

where D_t follows a K-state Markovian process with transition matrix

$$Pr(D_t = j|D_{t-1} = i) = p_{ij} \tag{4.20}$$

4.3 Inferences for the Correlation Matrix

Because of the high dimensionality and PSD constraints, estimation of a correlation matrix has been notoriously difficult. To complete our review, we summarize various approaches used in the above models for estimating correlation matrices.

4.3.1 VEC and BEKK Models

DCC-GARCH model resides in the VEC/BEKK model framework [17]. Given a correlation matrix which follows a GARCH type relationship as shown in Equation 4.7, a VEC version of the equation for Q_t is defined as:

$$vech(Q_t) = C^{\diamond}(\mathbf{1} - \mathbf{A}^{\diamond} - \mathbf{B}^{\diamond}) + \mathbf{A}^{\diamond} vech(\varepsilon_{\mathbf{t-1}}\varepsilon'_{\mathbf{t-1}}) + \mathbf{B}^{\diamond} vech(\mathbf{Q_{t-1}}) \tag{4.21}$$

where *vech* denotes an operator which stacks the lower triangle elements in a $p * p$ matrix as a vector of length $p(p+1)/2$. The vectors $A^{\diamond}$, $B^{\diamond}$ and $C^{\diamond}$ are constant parameter vectors of dimension $p(p+1)/2$. It has been shown [2] that Q_t is positive semi-definite for all t provided $A^{\diamond}$, $B^{\diamond}$ and $C^{\diamond}$ and initial Q are positive semi-definite. In this approach, re-weighting after estimation of Q is required to ensure that the diagonal elements in the correlation matrix Γ are ones.

4.3.2 Choleski Decomposition

The Regime Switching Dynamic Correlation GARCH model [44] adopts Choleski Decomposition to ensure PSD of the correlation matrix.

In a trivariate example provided in [44], the Choleski Decomposition of $\Gamma = P_n P'_n$ gives

$$\begin{pmatrix} p_{1,1}^2 & p_{1,1}p_{2,1} & p_{1,1}p_{3,1} \\ p_{1,1}p_{2,1} & p_{2,1}^2 + p_{2,2}^2 & p_{2,1}p_{3,1} + p_{2,2}p_{3,2} \\ p_{1,1}p_{3,1} & p_{2,1}p_{3,1} + p_{2,2}p_{3,2} & p_{3,1}^2 + p_{3,2}^2 + p_{3,3}^2 \end{pmatrix}$$

No re-weighting of the correlation matrix is needed after estimation. However, they imposed the artificial constraint that the off-diagonal elements in the correlation matrix are positive.

4.3.3 Stochastic Correlation

The time varying correlation matrices Γ_t in the DCC-MSV model [1] are unknown and stochastic, making the estimation more challenging than estimating for the multivariate GARCH model. This estimation complexity is resolved by simulating the latent matrices Q_t from its full conditional distribution by using a Metropolis Hasting Algorithm, which is described in detail in the Appendix of [1].

4.4 Inferences for the Hidden Regime Process in the Multivariate Framework

4.4.1 Likelihood Based Method

Estimation of both the Regime Switching Dynamic Correlation GARCH Model [44] and the Markov Switching Dynamic Conditional Correlation GARCH Model [6] are based on the Quasi Maximum Likelihood method using a Hamilton filter, which is similar to the QML method in the univariate case.

To clarify the prevailing estimation approach, we sketch the modified Hamilton filter and smoother algorithm [31], which is applied in the Markov Switching Dynamic Conditional Correlation GARCH Model.

A filter which samples from $P(D_t = j|I^t)$ is implemented iteratively forward as follows:

Step 1: Given the filtered probability $P(D_{t-1} = i|I^{t-1})$, determine the joint probabilities $P(D_t = j, D_{t-1} = i|I^{t-1}) = P(D_t = j|D_{t-1} = i)P(D_{t-1} = i|I^{t-1})$;

Step 2: Evaluate the regime dependent likelihood $\boldsymbol{L}_t(y_t|H_t, D_t = j, D_{t-1} = i, I^{t-1})$

Step 3: Evaluate the likelihood for observation:

$$\log \boldsymbol{L}_t(Y_t|H_t, I^{t-1}) = \sum_{j=1}^{K}\sum_{i=1}^{K} \boldsymbol{L}_t(Y_t|H_t, D_t = j, D_{t-1} = i, I^{t-1}) \tag{4.22}$$

and

$$\log \boldsymbol{L}(Y_t, ..., Y_1) = \log \boldsymbol{L}(Y_{t-1}, ..., Y_1) + \log \boldsymbol{L}(Y_t|H_t, I^{t-1}) \tag{4.23}$$

Step 4: Update the joint probabilities

$$P(D_t = j, D_{t-1} = i|I^t) = \frac{\log \boldsymbol{L}(Y_t|H_t, D_t, D_{t-1}, I^{t-1})P(D_t = j, D_{t-1} = i|I^{t-1})}{\log \boldsymbol{L}(Y_t|H_t, I^{t-1})} \tag{4.24}$$

Step 5: Compute the filtered probability

$$P(D_t = j|I^t) = \sum_{i=1}^{K} P(D_t = j, D_{t-1} = i|I^t) \tag{4.25}$$

Step 6: Update the correlation matrix Q.

Step 7: Iterate the above steps from $t = 1$ to $t = N$.

A smoother which samples from $P(D_t = j|I^N)$ is then implemented iteratively backward from $t = N$ to $t = 1$ as follows:

$$P(D_t = j, D_{t+1} = m|I^N) = \frac{P(D_{t+1} = m|I^N)P(D_t = j|I^t)P(D_{t+1} = m|D_t = j)}{P(D_{t+1} = m|I^t)} \tag{4.26}$$

$$P(D_t = j|I^N) = \sum_{m=1}^{K} P(D_t = j, D_{t+1} = m|I^N). \tag{4.27}$$

4.4.2 Bayesian Method

Bayesian hierarchical modeling is used to estimate parameters in the factor stochastic volatility model with switching regimes [38]. In this model, each element in $\mathbf{h}_t$ follows a univariate regime switching stochastic volatility model which is driven by the same underlying regime process $\mathbf{D}$. Therefore, in analogy to the univariate case, $\mathbf{D}$ is filtered and smoothed by a discrete time Markovian filter based algorithm, which has been described in Section 2.5.2.

4.5 Motivation for Proposed Multivariate Model

We are in favor of the multivariate regime switching covariance model proposed in [44] because on one hand it captures the dynamic structure of correlation and on the other hand, it is easier to estimate. However, it should be noted that our experience shows that not only correlations, but also stochastic volatilities are regime switching which has been observed in our univariate stochastic volatility models. Due to computational concerns, no one has dealt with regime switching in both the volatilities and correlations in a unified model. We extend the existing set of models by introducing a model that allows one regime switching process for the volatilities of each time series and one regime switching process for the correlation. For now, the regime processes are assumed to be independent. However, empirically, more volatile regimes tend to be related with higher correlation and in the future, extension can be made to allow certain patterns and relationships in the regime switching processes of correlations and volatilities.

With respect to the estimation of correlation matrix, either re-weighting is required or constraints must be imposed in existing models. In addition, none of the existing models allow a researcher to express substantive prior information about the strength of correlations among a set of variables. For example, stock returns within high technology industry are more correlated than those within financial industry. If this is known beforehand, the model can be improved by providing different prior information for each subset of variables. In our approach, we propose a Bayesian hierarchical modeling strategy to estimate correlation, which provides un-constrained estimates up to a high dimensionality. Under this framework, correlation can be grouped according to similarities of correlations or by incorporating prior information in the future.

One practical challenge that is often overlooked is the challenge associated with missing data. A variety of methodologies have been proposed to address this problem, but none has been seriously considered in the multivariate dynamic covariance framework. We contribute to this area by providing a Bayesian missing value imputation strategy, which leverages the information extracted from the correlation, volatilities and hidden regime processes to recover missing values.

As with the univariate models, all existing regime switching multivariate models are based on the assumption of Markovian regime switching processes. Our continuous-time version modeling of the process $\mathbf{D}$ is more flexible in allowing for a non-Markovian structure, which is shown to be important for certain types of real data.

Chapter 5

Regime Switching Multivariate Covariance Model

Starting in this chapter, we consider multiple time series, introduce our multivariate regime switching covariance model and demonstrate its usefulness. The most related work with our model is the Regime Switching Dynamic Correlation (RSDC) model proposed by Pelletier (2006) [44]. In his model, correlation is assumed to be constant within a regime and vary from regime to regime. It does not suffer from the curse of dimensionality and produces efficient forecasts. Our model preserves these appealing properties and in addition, we do not have to constrain the off-diagonal elements of the correlations to be positive as in the RSDC model. Moreover, a reversible jump MCMC algorithm allows for a non-Markovian process for the hidden regime process. Furthermore, by our MCMC sampler, we can sample regime switching processes for the volatilities and the correlations simultaneously.

If we assume there are p assets, the basic multivariate model is:

$$Y_t = H_t^{1/2} U_t \tag{5.1}$$

where Y_t is a vector of log returns of the p assets and the variable U_t is drawn randomly from an i.i.d $N(0, I_p)$. The time varying covariance matrix H_t can be decomposed into

$$H_t = S_t \Gamma_t S_t \tag{5.2}$$

where $S_t = \text{diag}(\exp(h_t^i/2))$ is a diagonal matrix of the assets' time varying stan-

dard deviations. The process h_t^i is the log volatility for the i^{th} asset at time t, which we have presented in the univariate case. The matrix $\Gamma_t = \{\rho_{ijt}\}_{p*p}$ is the $p * p$ correlation matrix with elements ρ_{ijt}. In our model, both the standard deviations and the correlations vary over time in a regime switching manner. More specifically,

$$h_{t+1}^i = \mu_{D_t^{ih}} + \phi_{D_t^{ih}} * (h_t^i - \mu_{D_t^{ih}}) + \tau_{D_t^{ih}} \varepsilon_t, \quad where \;\; \varepsilon_t \sim N(0,1) \tag{5.3}$$

$$\Gamma_t = \Gamma_{D_t^c} \tag{5.4}$$

where $\mathbf{D}^{ih}$ is the hidden volatility regime process for the i^{th} asset and $\mathbf{D}^c$ is the hidden correlation regime process. We therefore have $p+1$ different regime processes for the standard deviations and the correlations respectively. We model the most general case where all the regime processes are distinct and independent. However, it is also possible to constrain some assets to share a hidden regime process or to allow for dependence among the processes, a point we will discuss in Chapter 8.

5.1 Modeling Regime Switching Correlation

Without loss of generality, we assume $S_t = I$ to simplify the illustration of our regime switching correlation model. Integrating the volatility part into our model is straightforward in the Bayesian framework, which we describe later.

Assume there are K_c number of correlation regimes. Similar to the univariate case, $\mathbf{D}^c$ is assumed to consist of jump chains and waiting times, and the parameters to describe $\mathbf{D}^c$ include $(\boldsymbol{\alpha^c}, \boldsymbol{\beta^c}, \mathbf{P}^c)$, where $\boldsymbol{\alpha^c}$ and $\boldsymbol{\beta^c}$ are K_c dimensional vector parameters for the waiting time gamma distribution. If $\boldsymbol{\alpha^c} = \mathbf{1}$, the hidden regime process for the correlations is Markovian. $\mathbf{P}^c$ is the transition matrix given a jump occurrs.

The likelihood function for this model is:

$$f(\mathbf{Y}|\mathbf{\Gamma}, \boldsymbol{\alpha^c}, \boldsymbol{\beta^c}, \mathbf{P}^c) = \int f(\mathbf{\Gamma}, \boldsymbol{\alpha^c}, \boldsymbol{\beta^c}, \mathbf{P}^c, \mathbf{D}^c|\mathbf{Y}) d\mathbf{D}^c \tag{5.5}$$

where $\mathbf{Y} = (Y_1, ..., Y_N)$ and $\mathbf{\Gamma} = (\Gamma_1, ..., \Gamma_{K_c})$.

Because the likelihood is not in a closed form and difficult to evaluate, we adopt

the parameter space augmentation strategy. The parameter set is augmented to be

$$\boldsymbol{\Theta}^c = (\boldsymbol{\Gamma}, \boldsymbol{\alpha^c}, \boldsymbol{\beta^c}, \mathbf{P}^c, \mathbf{D}^c).$$

We now describe the estimation of the augmented parameter set, block by block.

5.1.1 Inference for the Correlation Regime Process

Conditional on $\mathbf{D}^c$, the inference for the parameters $(\boldsymbol{\alpha^c}, \boldsymbol{\beta^c}, \mathbf{P}^c)$ is exactly the same as in the univariate situation which is detailed in chapter 3.

The MCMC sampler for the hidden regime process $\mathbf{D}^c$ is also analogous to the algorithm described in the univariate stochastic volatility model. The only difference is that the calculation of the acceptance ratio ψ in each reversible jump MCMC sweep is based on the correlation matrix instead of the log volatilities in the univariate case. To be more specific,

$$\begin{aligned} \psi &= \frac{p(\mathbf{D}^c_{new}|-)q(\mathbf{D}^c_{old}|\mathbf{D}^c_{new})}{p(\mathbf{D}^c_{old}|-)q(\mathbf{D}^c_{new}|\mathbf{D}^c_{old})}, \\ p(\mathbf{D}^c_{new}|-) &\propto N(\mathbf{Y}|\boldsymbol{\Gamma}, \mathbf{D}^c_{new}) f(\mathbf{D}^c_{new}|\boldsymbol{\alpha^c}, \boldsymbol{\beta^c}, \mathbf{P}^c) \end{aligned} \tag{5.6}$$

where $p(\mathbf{D}^c_{new}|-)$ is the full conditional distribution of $\mathbf{D}^c_{new}$.

5.1.2 Inference for the Correlation Matrix

Conditional on the hidden regime process $\mathbf{D}^c$, we can update the correlation matrices $\Gamma_i, i = 1, ..., K_c$ in each MCMC sweep by a shrinkage analysis which is proposed in [35]. For clarity of illustration, we omit the subscript k and describe a general estimation method for a correlation matrix Γ.

We denote the (i, j) element in Γ as ρ_{ij} and we estimate Γ in a hierarchical structure.

Hierarchy 1

Assume each element ρ_{ij} in Γ follows a normal prior distribution $N(\delta_{ij}, \nu^2)$, where $\Delta = \{\delta_{ij}\} p * p$ is a layer of latent variables introduced in [35]. These latent variables are called "Shadow Priors" and are introduced to minimize computational problems associated with the PSD constraints in the correlation estimation. The

parameter ν^2 is a tuning parameter. The prior conditional density of Γ is

$$f(\Gamma|\Delta) \propto C(\Delta, \nu^2) \prod_{i<j} \exp -(\rho_{ij} - \delta_{ij})^2/(2\nu^2) I(\Gamma \in R^p) \tag{5.7}$$

where R^p is the space of all correlation matrices of dimension p and $C(\Delta, \nu^2)$ is the normalizing constant which equals

$$C(\Delta, \nu^2) = \int_{\Gamma \in R^p} \prod_{i<j} \exp -(\rho_{ij} - \delta_{ij})^2/(2\nu^2) d\rho_{ij} \tag{5.8}$$

Hierarchy 2

Each element δ_{ij} in the shadow prior matrix Δ is assumed to follow a common normal prior distribution, that is

$$f(\Delta|\widetilde{\mu}, \widetilde{\sigma}^2) \propto \prod_{i<j} \exp -(\delta_{ij} - \widetilde{\mu})^2/(2\widetilde{\sigma}^2) \tag{5.9}$$

where $(\tilde{\mu}, \tilde{\sigma})$ are the common mean and common variance hyper parameters.

Hierarchy 3

Diffuse priors are imposed and the hyper parameters in hierarchy 2 follow $\tilde{\mu} \sim N(0, \upsilon^2)$ and $\tilde{\sigma}^2 \sim IG(\bar{\alpha}, \bar{\beta})$.

Inference for Γ

Conditional on all the other parameters, the inference for Γ only involves the first hierarchy. The full conditional distribution for ρ_{ij} is

$$f(\rho_{ij}|-) \propto |\Gamma|^{-N/2} \exp(-tr(\Gamma^{-1}B)/2) \exp(-(\rho_{ij} - \delta_{ij})^2/(2\nu^2)) I(\Gamma \in R^p) \tag{5.10}$$

where B is the empirical correlation matrix. Since ρ_{ij} is embedded in the PSD constraints $I(\Gamma \in R^p)$, we sample ρ_{ij} one at a time using a Metropolis-Hasting algorithm [35]. The PSD property constrains ρ_{ij} to be on an interval (lb_{ij}, up_{ij}) and we propose ρ'_{ij} from this interval by a gridding uniform algorithm. The associated proposal probability is denoted by $p(\rho'_{ij})$.

The acceptance probability ψ

$$\psi = p(\rho'_{ij}) \frac{f(\rho'_{ij}|-)}{f(\rho^c_{ij}|-)} \tag{5.11}$$

where ρ_{ij}^c is the current correlation element and ρ_{ij}' denotes the proposed one.

Inference for Δ

The full conditional distribution for δ_{ij} is

$$\begin{aligned} f(\delta_{ij}|-) &\propto C(\Delta,\nu^2)\exp(-(\rho_{ij}-\delta_{ij})^2/(2\nu^2))\exp -(\delta_{ij}-\widetilde{\mu})^2/(2\widetilde{\sigma}^2), \\ &\propto C(\Delta,\nu^2)N(\mu^*,\sigma^{2*}) \end{aligned} \tag{5.12}$$

where $\mu^* = (\frac{\rho_{ij}}{\nu^2}+\frac{\widetilde{\mu}}{\widetilde{\sigma}^2})\sigma^{2*}$ and $\sigma^{2*} = \frac{\nu^2\widetilde{\sigma}^2}{\nu^2+\widetilde{\sigma}^2}$

Because Δ is embedded in $C(\Delta,\nu^2)$, we use a Metropolis Hasting Algorithm to sample δ_{ij} one at a time. Given the current value δ_{ij}^c, we propose δ_{ij}' from $N(\mu^*,\sigma^{2*})$ and accept this proposal with a probability $C(\delta_{ij}',\nu^2)/C(\delta_{ij}^c,\nu^2)$.

Because ν^2 is the tuning parameter over which we have complete control, in practice, we set ν^2 to a small value so that $C(\Delta',\nu^2)/C(\Delta^c,\nu^2)$ is approximately one. In this way we can avoid the cumbersome calculation of the constant C.

Inference for $\widetilde{\mu}$, $\widetilde{\sigma}^2$

By introducing the "Shadow Prior", the full conditional distributions of $\widetilde{\mu}$ and $\widetilde{\sigma}^2$ become conjugate and is easy to sample from:

$$\widetilde{\mu} \sim N(\frac{\upsilon^2\sum_{i<j}\delta_{ij}}{\frac{N(N-1)}{2}\upsilon^2+\widetilde{\sigma}^2}, \frac{\widetilde{\sigma}^2\upsilon^2}{\frac{N(N-1)}{2}\upsilon^2+\widetilde{\sigma}^2}) \tag{5.13}$$

$$\widetilde{\sigma}^2 \sim IG(\frac{N(N-1)}{4}+\bar{\alpha}-2, \frac{\sum_{i<j}(\delta_{ij}-\widetilde{\mu})^2}{2}+\bar{\beta}) \tag{5.14}$$

5.1.3 Tempering to Improve Mixing

Similar to the univariate estimation procedures, we allow for several MCMC chains to run in parallel with increasing temperatures to improve the mixing of parameters.

5.2 Modeling Regime Switching Stochastic Volatility and Correlation Simultaneously

Although envisioned in the paper of Pelletier [44], the situation where both the volatilities and the correlations are driven by a respective regime switching processes has never been documented in the literature. The explanation of such a vacancy is that such a model introduces too many switching parameters, which makes estimation unstable [44]. However in our Bayesian MCMC framework, integrating the univariate regime switching stochastic volatility model into the correlation model is straightforward.

The augmented parameters for the multivariate regime switching covariance model 4.1 is

$$(\boldsymbol{\Theta}^{h1},...,\boldsymbol{\Theta}^{hp},\boldsymbol{\Theta}^{c}) = ((\mathbf{h}^{i},\mathbf{D}^{hi},\boldsymbol{\theta}^{hi},\boldsymbol{\alpha}^{hi},\boldsymbol{\beta}^{hi},\mathbf{P}^{hi}), i=1,...,p, \quad (\boldsymbol{\Gamma},\boldsymbol{\alpha}^{\boldsymbol{c}},\boldsymbol{\beta}^{\boldsymbol{c}},\mathbf{P}^{c},\mathbf{D}^{c})) \tag{5.15}$$

As mentioned in Chapter 4, the log likelihood function of model 4.1 can be decomposed into two parts: a part for the standard deviations L_1, and conditional on the standard deviations, a part for correlations L_2. The equations are shown below:

$$\begin{aligned} L &= -\frac{1}{2}\sum_{t=1}^{N}(p\cdot\log(2\pi)+\log(|H_t|)+Y_t^{'}|H_t^{-1}|Y_t), \\ &= -\frac{1}{2}\sum_{t=1}^{N}(p\cdot\log(2\pi)+\log(|S_t\Gamma_t S_t|)+Y_t^{'}S_t^{-1}\Gamma_t S_t^{-1}Y_t), \\ &= -\frac{1}{2}\sum_{t=1}^{N}(p\cdot\log(2\pi)+2\log(|S_t|)+log(|\Gamma_t|)+\widetilde{U}_t^{'}\Gamma_t\widetilde{U}_t). \end{aligned} \tag{5.16}$$

where $\widetilde{U}_t = S_t^{-1}Y_t = (\widetilde{u}_{1.t},...,\widetilde{u}_{K,t})^{'}$ is a zero mean process with covariance matrix Γ_t. We can clearly identify two parts from the last equation, where

$$L_1 \propto 2\log(|S_t|) \tag{5.17}$$

Algorithm 6 MCMC Algorithm for Regime Switching Covariance Model

1: **Initialize** the volatilities of each asset using a univariate regime switching stochastic volatility model
2: **Start** the regime switching correlation model based on the standardized observations $y_t^{i*} = y_t^i / \exp(h_t^i/2)$ where $\mathbf{h}^i$ is obtained in the first step
3: **Update** volatilities and correlations and their regime processes simultaneously by MCMC sampling method.
   In **each sweep** of the MCMC sampling:
   1. Update all the parameters $\mathbf{\Theta}$ for each asset;
   2. Conditional on the updated volatilities $\mathbf{h}^i$,$i = 1, ..., p$, update all the correlation parameters $\mathbf{\Theta}^c$.

   Repeat the two steps until convergence

and conditional on the standard deviations,

$$L_2 \propto \log(|\Gamma_t|) + \widetilde{U}_t' \Gamma_t \widetilde{U}_t. \tag{5.18}$$

Therefore, to update volatilities and correlations simultaneously, we can simply estimate the univariate regime switching stochastic volatility model for each asset separately. Then conditional on the updated volatilities, estimate the regime switching correlation model. The resulting samples are drawn from the joint distribution of the augmented parameters in Equation 5.15. The algorithm is detailed in Algorithm 6. Extensive simulation and application to real data have demonstrated the efficacy of our model.

Chapter 6

Missing Data Imputation and Simulation

This chapter has three objectives: first, we propose and derive a Bayesian missing data imputation strategy; second, we present our simulation results with the purpose to: (1) provide evidence for the validity of the multivariate regime switching covariance model described in Chapter 5; (2) show the value of our imputation method with respect to correlation parameter estimation; third, we introduce and show the value of a useful application of our model: a dynamic portfolio allocation strategy.

6.1 Missing Data Imputation Method

Most existing models investigating the correlations among international stock markets are applied to weekly returns data. One reason for avoiding daily data is that the stock markets from different countries have different holiday schedules. For example, the US stock market is closed on Independence Day while in the other countries, stock markets remain open. In our experience, when we line up indices across countries according to calender days, around 4% of data is missing. This amount of missing data is not negligible and impacts model inferences and conclusions. An intuitive and common approach to solve the problem is to plug in the value from the previous day or the average value from the previous and following days and use the augmented data set to calculate either volatilities or correlations.

However, these plug-in methods ignore underlying information implied from historic data structure. We break from this initial approach and propose a Bayesian imputation approach which relies on all related data points and the model structure. While we illustrate our approach using examples from international equity markets, our method is general and can be applied to any multivariate missing data problem.

Assume we have daily indices data from p different countries of length $N+1$, each index's original value vector is denoted by $\mathbf{y}^{k*}, \quad k = 1, ..., p$ and it has a length of $N+1$. The log returns of each series are denoted by $\mathbf{y}^k$, where $y_t^k = log(y_{t+1}^{k*}) - log(y_t^{k*})$ and $\mathbf{y}^k$ has a length of N. Let $\tilde{\mathbf{y}}_t = (y_t^{1*}, ..., y_t^{p*}), t = 1, ..., N+1$ denote the original indices across p countries at time t. The derivation of the full conditional distribution of $\log(\tilde{\mathbf{y}}_t)$ is

$$f(\log \tilde{\mathbf{y}}_t|-) \propto f(\log \tilde{\mathbf{y}}_{t+1} - \log \tilde{\mathbf{y}}_t|h_t, \Gamma_{D_t^c}) * f(\log \tilde{\mathbf{y}}_t - \log \tilde{\mathbf{y}}_{t-1}|h_{t-1}, \Gamma_{D_{t-1}^c}) \quad (6.1)$$

where $\log \tilde{\mathbf{y}}_{t+1} - \log \tilde{\mathbf{y}}_t = \mathbf{Y}_t$ is the vector of log returns at time t. Therefore each part on the right hand side is distributed as

$$\log \tilde{\mathbf{y}}_{t+1} - \log \tilde{\mathbf{y}}_t|h_t, \Gamma_{D_t^c} \sim N(\mathbf{0}, H_t = S_t \Gamma_{D_t^c} S_t) \quad (6.2)$$

where $S_t = \text{diag}(\exp \frac{\mathbf{h}_t}{2})$.

Combining Equation 6.1 and Equation 6.3, we derive

$$f(\log \tilde{\mathbf{y}}_t|-) \propto N((H_t^{-1} + H_{t-1}^{-1})^{-1}(\log \tilde{\mathbf{y}}_{t+1} H_{t-1}^{-1} + \log \tilde{\mathbf{y}}_{t-1} H_t^{-1}), (H_t^{-1} + H_{t-1}^{-1})^{-1}) \quad (6.3)$$

As we have noted, certain portion of the data are missing for each series due to market closures. This means, that for each $\log \tilde{\mathbf{y}}_t, \ t = 1, ..., N+1$, there can be $0, 1, ..., p-1$ missing elements.[1] We denote the missing part as $\mathbf{y}_t^-$ and the remaining part as $\mathbf{y}_t^+$. Based on the classical conditional multivariate normal distribution equation [2], we can derive the full conditional distribution for the missing portion $f(\log \mathbf{y}_t^-|-)$ and easily draw samples from it. It is straightforward to integrate the sampler for missing data into each sweep of our MCMC algorithm. This approach

[1] Note that we delete the data entry where no international market is open on that day.

[2] http://en.wikipedia.org/wiki/Multivariate_normal_distribution

allows us to extract information from and at the same time influence the inference of the dynamic covariance structures.

6.2 Simulation Study

In this section we provide evidence of the validity and effectiveness of our model through a simulation study. Our model accurately reproduces parameters and recovers both volatility and correlation regime switching processes and provide estimators that are stable over different initializations. The proposed Bayesian method to treat missing data is also shown to outperform other imputation methods. Finally, we introduce a dynamic portfolio allocation strategy based on our model, which reduces cost and produces higher Sharpe Ratios than other popular strategies in the literature.

We simulate log return data with four assets ($p = 4$). Each asset has a length of $N = 4000$. The volatilities for each asset follow a 2-regime switching stochastic volatility model, which is the same as in the simulation study of the univariate models. We specify $\mu_1 < \mu_2$, $\phi_1 = \phi_2$ and $\tau_1^2 = \tau_2^2$ and two regimes are also assumed for the correlations. Starting from a random initialization, we sample 50,000 burn-in sweeps and 100,000 sampling sweeps, with tempered chains for both of the correlations and the volatilities.

The simulated series $\mathbf{y}^i,\ i = 1, ..., 4$ and the posterior probabilities for the volatility regimes are shown in Figure 6.1. Each panel in this figure presents one set of data. The first plot in each panel shows the generated log-return series and the posterior probabilities for regime one and two are plotted below it. Clearly low volatility periods of log-returns correspond to a posterior probability of almost one for being in regime one and a high posterior probability, near one for regime two, corresponds to high volatility periods. These results show that we can successfully identify the volatility regimes for each series using our MCMC algorithm. To simplify the discussion, we do not present the volatility results here; interested readers can refer to Chapter 3 for more information.

The simulated regime process for the correlations and the posterior probabilities of each regime are shown in Figure 6.2. This figure suggests that our algorithm is able to recover the correlation regimes with high precision. Moreover, our al-

gorithm is stable over different random initializations, which answers the question raised by Pelletier about the possibility of implementing models with both regime switching volatilities and regime switching correlations [44].

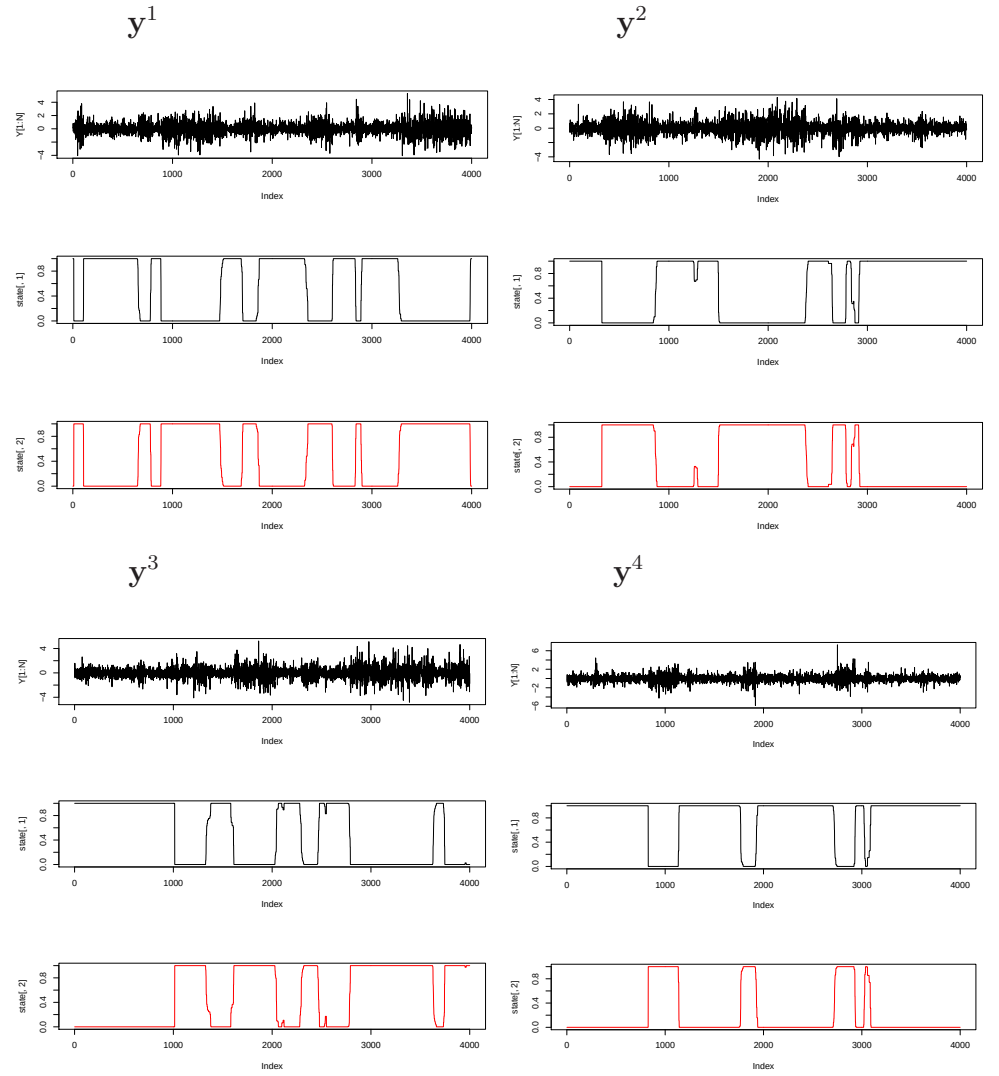

Figure 6.1. Simulated data for four time series, y1-y4. In each panel, we plot the returns Y and the posterior probability of being in each volatility regime which was estimated by our dynamic covariance model

To illustrate the advantage of the Bayesian imputation method for missing data, we randomly remove 4% data for each asset. The percentage of missing data is selected to mimic international equity index data. Three different missing

value recovery strategies are tested and compared: a Bayesian imputation method, the method that plugs in the previous index value, and the method that plugs in the average value of the previous and next indices. We have observed in previous chapters that with every single MCMC initialization our algorithm produces accurate estimators and well shaped posterior parameter distributions (Table 3.2, Figure 3.3). While we continue to obtain well behaved MCMC estimates, we want to investigate to get a sense of the uncertainties in the parameter posterior estimators over different initializations. We want to ensure that, in practice, our implementation of the full MCMC algorithm (Algorithm 6) is not trapped in a local maximum, or that different initializations lead to different estimators. As a result, in addition to incorporating parallel tempering to improve MCMC mixing, we run the full MCMC algorithm 10 times with different initializations for each imputation strategy on the missing data. Meanwhile, the same ten initialized MCMC algorithm is applied on the complete data.

The estimated correlation matrices are shown in Table 6.1 for regime one and in Table 6.2 for regime two. There are five panels in each table: the first panel provides the generating empirical correlation matrix, which we call "true matrix", or the empirical correlation for each regime based on true $\mathbf{D}^c$. The second panel presents the posterior means and standard deviations (in the first parenthesis) for the correlation computed with the complete data set. The following three panels report the posterior correlation estimators computed with missing data using different imputation strategies. Besides the typical posterior standard deviation estimators shown in the first parenthesis, we provide another type of standard deviations (we name them "type II standard deviation") for the correlation elements in the second parenthesis of each cell in the table. The type II standard deviations are the standard deviations of parameter posterior means over the 10 initializations and they provide evidence of model stability with random start. The insights from the tables are twofold: first, comparing the correlations between the first and second panel, we conclude that our MCMC algorithm can accurately recover each element of the correlation matrices as well as the hyper mean parameter $\tilde{\mu}$ and the estimators are stable because we have small type II standard deviations. Second, our proposed Bayesian imputation method outperforms the other two popular data imputation strategies because the estimates using our proposed imputation

method are closer to the truth.

To summarize, our model can recover volatility parameters, correlation matrices, and at the same time, the regime switching processes for the volatilities and the correlations. As for the concern of model stability [44] by introducing a large number of switching parameters, the small type II standard deviations of parameter posterior means in Table 6.1 and Table 6.2 suggest that our estimation algorithm reaches the global maximum and is stable. The differences between the posterior and the true correlations most likely rise from the shrinkage part of the correlation model, where all the elements in the correlation matrix shrink, or are "dragged" towards their common mean.

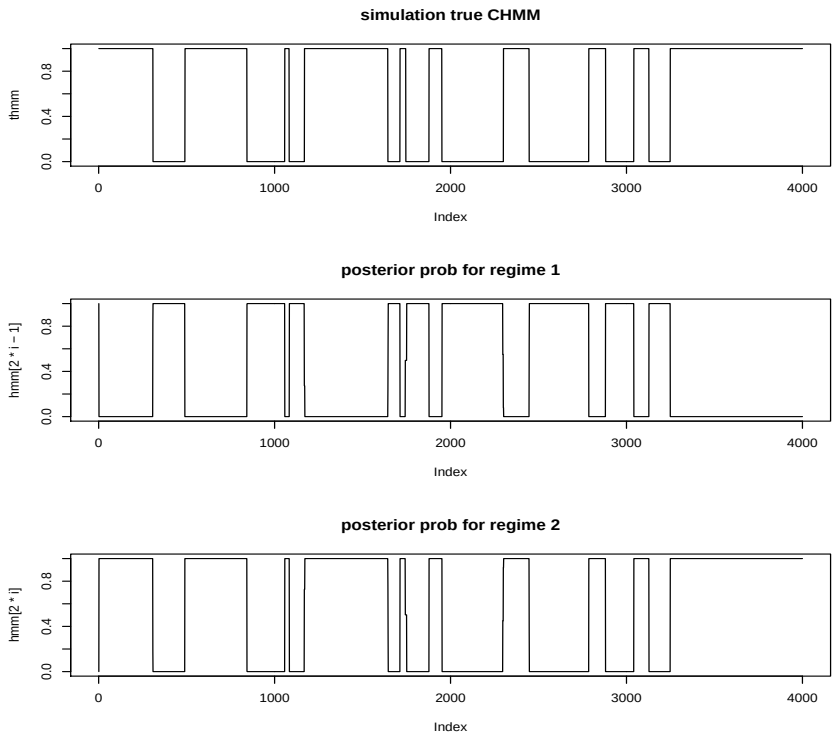

Figure 6.2. Posterior regime probabilities for simulated multivariate data: the first plot shows the true correlation regimes; the second shows the posterior probabilities of being in correlation regime 1 and the third shows the posterior probabilities of being in correlation regime 2

Regime 1			
True	$\tilde{\mu} = -0.271$		
1	-0.488	-0.165	-0.059
*	1	-0.469	-0.224
*	*	1	-0.235
*	*	*	1
No Missing	$\tilde{\mu} =$	-0.262(0.02)(0.002)	
1	-0.459(0.03)(0.004)	-0.168(0.02)(0.004)	-0.053(0.02)(0.001)
*	1	-0.467(0.02)(0.003)	-0.198(0.02)(0.004)
*	*	1	-0.234(0.02)(0.003)
*	*	*	1
Bayesian	$\tilde{\mu} =$	-0.257(0.02)(0.001)	
1	-0.450(0.02)(0.001)	-0.162(0.02)(0.003)	-0.047(0.02)(0.001)
*	1	-0.454(0.02)(0.002)	-0.199(0.01)(0.004)
*	*	1	-0.226(0.02)(0.004)
*	*	*	1
Average	$\tilde{\mu} =$	-0.247(0.02)(0.001)	
1	-0.436(0.03)(0.004)	-0.159(0.02)(0.004)	-0.042(0.02)(0.001)
*	1	-0.446(0.02)(0.002)	-0.196(0.02)(0.003)
*	*	1	-0.217(0.01)(0.004)
*	*	*	1
Previous	$\tilde{\mu} =$	-0.224(0.02)(0.003)	
1	-0.384(0.02)(0.004)	-0.121(0.02)(0.003)	-0.044(0.02)(0.002)
*	1	-0.407(0.03)(0.003)	-0.179(0.02)(0.003)
*	*	1	-0.202(0.02)(0.003)
*	*	*	1

Table 6.1. Posterior means, standard deviations (in the first parenthesis) and the type II standard deviations (in the second parenthesis) of correlation of regime 1 for simulation data: we compare different missing value imputation strategies

6.3 Model Comparisons

Using the simulated data, we assess the performances of various competing models. The competing models include:

- Regime switch in both the correlation and the volatilities (CV Switch);
- Regime switch in the correlation but standard, non-regime switching stochastic volatility model (C Switch);

Regime 2			
True	$\tilde{\mu} = 0.717$		
1	0.763	0.845	0.443
*	1	0.816	0.725
*	*	1	0.711
*	*	*	1
No Missing	$\tilde{\mu} =$	0.712(0.01)(0.001)	
1	0.739(0.01)(0.001)	0.827(0.01)(0.002)	0.441(0.01)(0.002)
*	1	0.799(0.01)(0.001)	0.711(0.01)(0.001)
*	*	1	0.703(0.01)(0.001)
*	*	*	1
Bayesian	$\tilde{\mu} =$	0.692(0.01)(0.001)	
1	0.725(0.01)(0.001)	0.815(0.01)(0.001)	0.428(0.01)(0.002)
*	1	0.785(0.01)(0.001)	0.689(0.01)(0.001)
*	*	1	0.687(0.01)(0.002)
*	*	*	1
Average	$\tilde{\mu} =$	0.666(0.01)(0.001)	
1	0.698(0.01)(0.001)	0.783(0.01)(0.002)	0.413(0.01)(0.001)
*	1	0.757(0.01)(0.001)	0.666(0.01)(0.001)
*	*	1	0.664(0.01)(0.001)
*	*	*	1
Previous	$\tilde{\mu} =$	0.592(0.01)(0.001)	
1	0.603(0.01)(0.001)	0.679(0.01)(0.002)	0.359(0.01)(0.002)
*	1	0.676(0.01)(0.001)	0.610(0.01)(0.001)
*	*	1	0.603(0.01)(0.001)
*	*	*	1

Table 6.2. Posterior means, standard deviations (in the first parenthesis) and the type II standard deviations (in the second parenthesis) of correlation of regime 2 for simulation data: we compare different missing value imputation strategies

- Regime switch in the stochastic volatilities but constant correlation (V Switch);
- Constant correlation and standard, non-regime switching stochastic volatility model (No Switch);

All the four models mentioned above are applied on the simulated complete data set. The models are also applied in combination with different imputation strategies on the multivariate data set with 4% of values missing.

We use a subset of the model comparison tools introduced in Section 3.8, which include the harmonic mean based marginal likelihood and DIC. In addition, we also compare the hyper-mean parameter $\tilde{\mu}_i$ estimator for each regime. Finally, we test the performance of each model, in a practical setting, by using the parameter estimates in an asset allocation strategy, which we introduce later.

6.3.1 Model Comparison Results

The model comparison results from the simulated data are presented in Table 6.3. Panel (a) presents the marginal likelihoods calculated by harmonic mean and panel (b) presents the results of the DIC method. The first row in each panel shows the results with complete data. The second to the fourth rows correspond to simulated missing data, with Bayesian imputation method (Bayes), imputation with the average between the previous and following values (Average) and imputation with previous value (Previous) respectively. Each column presents the results for one of the competing models, where the models were specified at the beginning of Section 6.3 and the boldfaced column corresponds to our proposed model. Clearly our proposed full model always produces larger marginal likelihoods and smaller DIC values compared to other models, which demonstrates a better fit.

(a)

Harmonic Mean	**CV Switch**	C Switch	V Switch	No Switch
All Data	**-15591(4)**	-16171(3)	-19414(4)	-19690(4)
Bayes	**-15885(9)**	-16479(14)	-19464(11)	-19750(12)
Average	**-15863(4)**	-16383(4)	-18994(4)	-19263(6)
Previous	**-16282(3)**	-17348(3)	-18514(4)	-19859(6)

(b)

DIC				
All Data	**31292(73)**	32387(67)	38707(82)	39479(100)
Bayes	**31763(85)**	33040(90)	38594(100)	39378(106)
Average	**31843(58)**	32824(56)	37851(61)	38579(63)
Previous	**32499(54)**	34682(56)	37020(64)	39687(74)

Table 6.3. Likelihood based model comparison results for different models (column) with the simulated multivariate data: (a) marginal likelihood calculated by harmonic mean; (b) DIC values. The numbers in the parenthesis are the type II standard deviations of the model comparison metrics from 10 differently initialized MCMC algorithms.

	CV Switch	C Switch	V Switch	No Switch
Regime 1				
Truth	$\tilde{\mu} = -0.271$			
All Data	**-0.262(0.02)**	-0.255(0.02)	0.302(0.02)	0.299(0.02)
Bayes	**-0.257(0.02)**	-0.248(0.02)	0.296(0.02)	0.294(0.02)
Average	-0.247(0.02)	-0.245(0.02)	0.288(0.02)	0.280(0.02)
Previous	-0.224(0.02)	-0.231(0.02)	0.252(0.02)	0.249(0.02)
Regime 2				
Truth	$\tilde{\mu} = 0.717$			
All Data	**0.712(0.01)**	0.691(0.01)	*	*
Bayes	**0.692(0.01)**	0.678(0.01)	*	*
Average	0.666(0.01)	0.657(0.01)	*	*
Previous	0.592(0.01)	0.610(0.01)	*	*

Table 6.4. Posterior means and standard deviations of hyper mean $\tilde{\boldsymbol{\mu}}$ estimations for simulation data: we compare different models (by columns) and different missing value imputation strategies (by rows).

The parameter estimation based model comparison results are shown in Table 6.4. The layout of this table is similar to Table 6.3. To save space, we only present the comparisons with respect to hyper-mean $\tilde{\boldsymbol{\mu}}$. Reader can refer to Table 6.1 and Table 6.2 for detailed correlation estimation results using our proposed full model.

Table 6.4 offers two important insights. First, the Bayesian imputation method is preferred. Among all the competing models, it holds in general that the estimated hyper means $\tilde{\boldsymbol{\mu}}$ computed from the Bayesian imputation method are the closest to those with the complete data set. This supports the claim that the Bayesian imputed missing values are the closest to the true values. Second, the regime switching covariance model outperforms the other models. To find the best performing model, we should compare different models within rows. For each row in Table 6.4, our proposed full model always outperforms all the other models with regards to the most accurate estimation of parameters, where the true values are $\tilde{\mu}_1 = -0.271$ and $\tilde{\mu}_2 = 0.717$. This result is true for both the complete data and with any imputation method for the missing data.

6.4 Dynamic Portfolio Allocation using Multivariate Regime Switching Covariance Model

A potential valuable use of multivariate covariance modeling is in determining an optimal portfolio allocation. The objective of a portfolio manager is to maximize the return of a portfolio while controlling for risks. The Sharpe Ratio is a metric that is often used to assess the performance of a portfolio. It measures the excess return per unit of risk for a portfolio and a larger Sharpe Ratio is preferable. Without loss of generality, we assume the risk free rate is zero. Therefore a Sharpe Ratio is defined as $\frac{R_p}{\sigma_p}$, where $R_p = w'R$ and $\sigma_p = \sqrt{w'Hw}$ are the expected return and standard deviation for the portfolio respectively. The vector w is the weight vector, R is the mean return vector, and H is the covariance matrix of the assets in a portfolio. To quantify the objective, we want to minimize the following expression:

$$w'Hw - qw'R \tag{6.4}$$

with the constraint that $\sum_i w_i = 1$, where q is a risk aversion parameter.

According to the Markowitz Portfolio Theory [40], the optimal risky portfolio (assuming risk free rate is zero) has a weight:

$$w = \frac{H^{-1}R}{R'H^{-1}\mathbf{1}} \tag{6.5}$$

where $\mathbf{1}$ is the vector of 1s.

We do not address the time varying mean asset return case and assume the mean value is constant throughout our work. For the covariances, abundant empirical evidences indicate that it is not constant over time; therefore, a Sharpe Ratio calculated based on a constant covariance matrix assumption is not accurate. It is beneficial to adjust portfolio weights in a timely manner so as to capture the time varying covariance structures, and under the assumption of no transaction cost, the ideal strategy is to update weights daily, based on the return and covariance on that day. In practice, in order to avoid excessive transaction charges, portfolio managers adjust weights with some frequency lower than daily according to the up-to-date covariance information. In this section, based on our regime switching model, we propose an algorithm where portfolio weights are adjusted when either

Algorithm 7 Asset Allocation

1: Calculate average log return for each asset series R_i

2: At $t = 1$

The covariance matrix H is the posterior covariance computed from the model at $t = 1$.

Without constraining the weights to be positive, calculate optimal weights w according to Markowitz portfolio theory 6.5.

Calculate Sharpe Ratio according to the calculated weights w.

3: For t from 2 to N

a. KEEP the weights the same unless any of the following switch occurs:

A switch in the volatility of Asset 1;

.....;

A switch in the volatility of Asset p;

A switch in the correlation.

b. If the weights w stay the same, calculate Sharpe Ratio according to the weights and the posterior covariance matrix at time t;

c. If a switch occurs, recalculate weights according to the mean volatility $\mu_{D_t^{ih}}$ and the correlation $\Gamma_{D_t^c}$ for the new regime. Then calculate Sharpe Ratio according to the new weights.

Iterate step a, b, c.

4: Take average of the calculated Sharpe Ratio over time.

a volatility or the correlation regime switches states. To define a regime switch based on the posterior probabilities of a regime process, we create an artificial chain $\mathbf{D}^a$ where $D_t^a = L,\ L = \arg\max(p_t^L)$ and p_t^L denotes the posterior probability of being in regime L at time t. When $\mathbf{D}^a$ switches regime, it makes intuitive sense to re-balance the portfolio accordingly. This algorithm that implements this strategy is detailed in Algorithm 7.

For the "V Switch" model which assumes a constant correlation, the weights are re-calculated only when the mean volatility levels change. While for the "C Switch" model where there is no regime switching specified for the volatility, the weights are re-calculated when the correlations switch regimes. The number of times we re-balance a portfolio is counted and stored. For a fair comparison, we also update portfolio weights at regular intervals, for example, every 40 days, so that the number of updates is comparable to our regime switching portfolio adjustment strategy. The model and rebalancing strategy which provides a higher Sharpe Ratio conditional on the same number of portfolio re-balances is desirable.

6.4.1 Portfolio Allocation Results

The portfolio allocation results for the simulated data are shown in Table 6.5. There are three panels in this table, each corresponding to a portfolio re-balance strategy. We have (a) daily re-balance; (b) regular re-balance; (c) regime switch re-balance. Each row in this table shows the results from one model, and the cells within a row share the same number of re-balance times. For example, all the cells in panel (b) correspond to 92 re-balance times while the rows in panel (c) correspond to 91, 20, 75 re-balance times respectively. As a result we compare Sharpe Ratio within rows. The columns correspond to different data set or imputation method. The value 0.176 in the "True" column is the Sharpe Ratio calculated based on true model parameters and a daily re-balance strategy, and not surprisingly as it is the ideal case, it produces the highest Sharpe Ratio. We use this case as a comparison benchmark. The "All data" column shows the results for simulated data where none of the data is missing and all the other columns show the results for simulated data with some data missing and with different imputation methods.

Table 6.5 offers three conclusions: The first two are the same as the conclusions from Table 6.4. The Bayesian imputation method generates higher Sharpe Ratios than other imputation methods, which demonstrates the superiority of the Bayesian imputation method. In order to compare the performances of different covariance models, we should only look at panel (a) and panel (b) where the re-balance numbers are the same within columns. It is easy to find that our regime switching covariance model (highlighted) produces higher Sharpe Ratios when compared to other models.

The third conclusion is with respect to the portfolio measurement perspective. It is easy to understand that updating portfolio weights daily produces the highest average Sharpe Ratio because we maximize the Sharpe Ratio for each day according to the up-to-date covariance matrix. From the "All Data" column within panel (a), we can see that with no missing data, the highest Sharpe Ratio is achieved by our model, which is 0.168. However updating portfolio daily is too expensive and not practical. Portfolio managers hence adopt a regular re-balance strategy, or to adjust portfolio weights once every period. For a fair comparison objective which we discuss later, we update weights every 44 days, which results in 92 adjustments.

Using this approach, the average Sharpe Ratio for "All Data" drops to 0.116 based on our model, which is shown in panel (b). The dramatic decrease in the Sharpe Ratio is not desirable. Therefore we propose the regime switching portfolio re-balance strategy in Algorithm 7 which re-balances a portfolio only when a switch in regime is detected. The results are shown in panel (c). With this strategy, we adjust portfolio weights 91 times, which is almost equivalent to the frequency in the regular re-balance case. However, the average Sharpe Ratio is 0.167, which is only 0.001 lower than the maximum value achieved by a daily update. Using the simulated data, our proposed re-balance strategy strikes a astonishingly good balance between achieving a large Sharpe Ratio and reducing re-balance frequency. This dramatic improvement in portfolio performance gives further evidence of the advantage of our proposed models and algorithms.

Portfolio Allocation

(a)

	True	All Data	Bayes	Average	Previous
Daily Re-balance					
VC Switch	**0.176**	**0.168**	**0.163**	**0.157**	**0.126**
C Switch	*	0.140	0.135	0.129	0.111
V Switch	*	0.113	0.111	0.110	0.102
Neither Switch	*	0.089	0.087	0.086	0.085
Constant	*	0.042	0.042	0.039	0.033

(b)

Re-balance Every 44 day (92)					
VC Switch	*	**0.116**	**0.116**	**0.115**	**0.086**
C Switch	*	0.099	0.099	0.097	0.093
V Switch	*	0.084	0.084	0.080	0.071
Neither Switch	*	0.073	0.077	0.071	0.071
Constant	*	0.028	0.025	0.022	0.02

(c)

Regime Switching Re-balance					
VC Switch(91)	*	**0.167**	**0.164**	**0.160**	**0.138**
C change(20)	*	0.010	0.009	0.008	0.006
V Switch(75)	*	0.114	0.114	0.116	0.112

Table 6.5. Portfolio allocation results for different models, column, with different re-balance strategies for the simulated multivariate data: (a) daily re-balance; (b) regular re-balance (c) regime switch re-balance

Chapter 7

Empirical Analysis

In this chapter, we apply our regime switching covariance model and Bayesian imputation algorithm to daily international indices data. We compare the empirical performances of our model with some alternative models. In addition, within our model framework, we show how to determine the number of regimes describing the dynamics of correlations and volatilities. The performances of different models are compared in two ways: the likelihood related criteria and the posterior probability for each model. We also apply the regime switching asset allocation strategy to this real data and illustrate the value of these models in a practical setting.

7.1 Data Description

The data we work with are weekday closing values for four countries: US (S&P 500), UK (FTSE), Germany (DAX 30) and Japan (NIKKEI 225). The sample goes from 1993/06/26 to 2008/9/28 and contains a total of 4000 observations for each country. The four series are lined up by calender days, which gives rise to approximately 4% missing data for each series because of non-overlapping holidays. Based on the conclusion from the simulation study in Chapter 6, we adopt the Bayesian imputation method for missing data throughout our empirical analysis. The log-returns for each series are de-meaned and magnified 100 times; in addition, we remove common holidays, e.g., new years day, where no market is open for any of the four countries. The descriptive statistics for each time series are presented in Table 7.1 and the empirical covariance and correlation matrices are provided in

Table 7.2. Among the four indices, Japan exhibits an undesirable behavior of a negative average log return and large standard deviation.

Because of the complexity of having multiple regimes for volatilities and correlations, we follow a two-step estimation procedure. In the first step, we model univariate stochastic volatilities for each asset following the algorithm introduced in Chapter 3. By comparing different models, we decide on the number of volatility regimes and whether we should allow $\boldsymbol{\phi}$ and $\boldsymbol{\tau}^2$ to differ across regimes. In the second step, the decisions about volatilities from step 1, such as the number of regimes for each univariate volatility process, are then incorporated into the full MCMC algorithm in step 2. We estimate dynamic covariance matrices according to Algorithm 6 and decide on the number of correlation regimes in the second step. For all the models we work with, we implement the MCMC procedure independently with 10 different initializations to test for model stability and we take average of the parameter means from each different initialization.

	US	UK	GERMANY	JAPAN
Mean	0.0253	0.015	0.0299	-0.0134
Standard Deviation	1.038	0.971	1.424	1.378
Median	0.027	0.031	0.0669	0
Skewness	-0.15	-0.286	-0.278	-0.059
Kurtosis	3.72	3.19	3.58	2.32
Minimum	-7.11	-5.41	-8.87	-7.23
Maximum	5.57	5.09	7.55	7.66

Table 7.1. Descriptive statistics of daily indices closing value log return series (times 100)

7.2 Stochastic Volatility Modeling

For each univariate time series, model comparisons [1] indicate that a two regime switching stochastic volatility model, where all the volatility parameters, $\boldsymbol{\mu}$, $\boldsymbol{\phi}$, $\boldsymbol{\tau^2}$, differ across regimes, is the best.

The log return data for each series, along with the posterior probabilities of volatility regimes are displayed in Figure 7.1. The US and UK indices have similar

[1] To save space, we do not present the results here.

Empirical Covariance

	US	UK	GERMANY	JAPAN
US	1.077	0.421	0.693	0.137
UK	*	0.943	1.006	0.351
GERMANY	*	*	2.028	0.447
JAPAN	*	*	*	1.901

Empirical Correlation

	US	UK	GERMANY	JAPAN
US	1.0	0.418	0.469	0.096
UK	*	1.0	0.727	0.262
GERMANY	*	*	1.0	0.227
JAPAN	*	*	*	1.0

Table 7.2. Empirical covariance and correlation matrices of daily indices log return series (times 100)

patterns in the volatilities as there are two regimes with infrequent switches. The volatility dynamics of the Germany and Japan indices are more complicated and show more frequent and irregular switches in the regimes. Because we do not have evidence of frequent regular switches as would arise in the non-Markovian regime switching for either the volatility or the correlation, we fit waiting times using an exponential distribution.

The volatility parameter estimates for each regime are shown in Table 7.3. As in Section 6.2, the numbers in parenthesis are standard deviations of posterior parameter estimators over 10 differently initialized MCMC samples. This definition of standard deviations in parenthesis holds throughout this chapter. Small standard deviations indicate that our parameter inferences are not sensitive to different random seed initialization and stable, which again addresses the concern of model instability in [44]. One interesting point for this data set is that a high volatility regime is related to a higher persistence and lower variance of the volatility. This observation has an econometric implication that during a volatile period of time, once an extreme event happens which drags volatility up from its mean level, it takes much longer time for the volatility to come back to the mean level because high persistence is associated with high volatility.

Asset	μ_1	μ_2	ϕ_1	ϕ_2	τ_1^2	τ_2^2
US	-1.12(0.02)	0.32(0.02)	0.59(0.06)	0.81(0.1)	0.29 (0.02)	0.13(0.01)
UK	-1.22(0.03)	-0.068(0.01)	0.34(0.07)	0.74(0.09)	0.26(0.01)	0.20(0.01)
GERM	-0.61(0.01)	0.61(0.02)	0.44(0.01)	0.77(0.07)	0.26(0.02)	0.18(0.01)
JAPAN	-0.41(0.01)	0.69(0.01)	0.44(0.04)	0.83(0.08)	0.39(0.02)	0.11(0.01)

Table 7.3. Posterior mean estimators of volatility parameters and the standard deviations of posterior mean estimators over different initializations (in parenthesis) for regime switching covariance model of international indices data, 1993-2008

7.3 Correlation Modeling

We fit models with anywhere from two to five regimes for the correlation and report the likelihood based model comparison results in Table 7.4. Model comparisons indicate that there are four correlation regimes in this data.

We compare the full four regime model with simplified versions listed in Section 6.3 and report the results in Table 7.5. The "CV switch" model refers to our proposed full model with two regimes and four regimes for the volatilities and the correlations respectively. The full model provides the largest marginal likelihood through both the harmonic mean and the filtering calculation while the DIC criterion favors the model of a constant correlation and switching stochastic volatilities (the "V switch" model). This difference could be due to the penalty of complex models imposed by the DIC criterion, as there are 24 more parameters in our model than the "V switch" model. Nevertheless, the difference in the DIC values between the full model and the "V switch" model is not significant and the proposed full model fits better than other models with respect to the model comparison criteria.

	2 Regime	3 Regime	4 Regime	5 Regime
HarmonicMean	-9203(7)	-9168(9)	**-9132(5)**	-9130(9)
Filter	-9388(30)	-9442(36)	**-9341(45)**	-9343(32)
DIC	-204895(21)	-204957(26)	**-205054(25)**	-205034(26))

Table 7.4. Determine the number of correlation regimes by model comparison methods for international indices data, 1993-2008. 4 Regime is selected. The numbers in the parenthesis are the type II standard deviations of the model comparison metrics from 10 differently initialized MCMC algorithms.

The posterior probabilities for the correlation regimes are presented in Figure

	CV switch	C switch	V switch	No switch	Const
Harm Mean	**-9132(5)**	-9272(17)	-9341(7)	-9435(7)	-12006(6)
Filter	**-9341(45)**	-9744(36)	-13341(45)	-13269(25)	*
DIC	**-205054(25)**	-205041(26)	-205440(20)	-183825(20)	-105725(47)

Table 7.5. Model comparisons based on 4 correlation regimes and 2 volatility regimes specification for International Indices data, 1993-2008. The numbers in the parenthesis are the type II standard deviations of the model comparison metrics from 10 differently initialized MCMC algorithms.

7.2. Regime one to four corresponds to the lowest to the highest average correlation regimes. This figure suggests that average correlations go up in jumps, or the correlations stay in one regime for a while and suddenly switch to a different, more highly correlated regime. Once the correlations leave the previous regime, they do not return to the earlier regime. Another implication from this figure is that the switches in the correlation regimes do have some correspondence with the switches in volatilities. For example, the indices start with a low volatility regime which also corresponds to lower correlations among the indices. Then possibly in combination with the occurrence of the Asian financial crisis around the end of 1997, the volatilities increase, and the financial markets from these countries become more correlated. This structure lasts till the 2001 terrorist attack, and after that, all indices (except Japan) experience an even more volatile period of time, which drags up correlations as well. Finally, for the most recent several years, possibly because of globalization, the indices are stable and highly correlated.

In order to better understand how the correlations change over time, a close scrutiny of each element of the correlation matrices is appropriate. The hyper mean, denoted by $\tilde{\mu}$, of the off-diagonal elements of each correlation matrix represents the average level of the correlations among the indices. From Table 7.6, we find that regime one has a significantly lower average correlation and the average correlation increases from the second to the fourth regime. These increases are not as sharp as the increases from regime one. While we call regime two through four the "high correlation regimes", the correlation structures are not identical within these "high correlation" regimes. In 2001, the correlation enters a regime where the correlations between the US and other indices increase while the one between Germany and Japan decreases. This structure is maintained for several years until

around the middle of 2004, when the correlation between Germany and the US drops while the correlations between Germany and UK and Japan increase. In the final regime, the correlation between Germany and UK increases to 0.8, which indicates a sudden increased closeness between these two countries.

Regime 1 $\tilde{\mu} = 0.219$	US	UK	GERMANY	JAPAN
US	1	0.30(0.01)	0.14(0.02)	0.05(0.01)
UK	*	1	0.43(0.03)	0.15(0.01)
GERMANY	*	*	1	0.24(0.01)
JAPAN	*	*	*	1
Regime 2 $\tilde{\mu} = 0.343$	US	UK	GERMANY	JAPAN
US	1	0.38(0.01)	0.35(0.02)	0.13(0.01)
UK	*	1	0.64(0.03)	0.28(0.01)
GERMANY	*	*	1	0.27(0.01)
JAPAN	*	*	*	1
Regime 3 $\tilde{\mu} = 0.379$	US	UK	GERMANY	JAPAN
US	1	0.42(0.03)	0.58(0.03)	0.15(0.02)
UK	*	1	0.66(0.02)	0.27(0.01)
GERMANY	*	*	1	0.19(0.02)
JAPAN	*	*	*	1
Regime 4 $\tilde{\mu} = 0.408$	US	UK	GERMANY	JAPAN
US	1	0.43(0.02)	0.47(0.01)	0.10(0.02)
UK	*	1	0.80(0.02)	0.34(0.01)
GERMANY	*	*	1	0.33(0.01)
JAPAN	*	*	*	1

Table 7.6. Posterior mean estimators of correlations and the type II standard deviations of posterior means over different initializations (in parenthesis) of each regime for international indices data, 1993-2008

The phenomenon that each regime or correlation structure appears only (Figure 7.2) once does not always hold. For example, when we consider the same four

indices from 1980/1/1 to 1991/07/01, the posterior probabilities for the correlation regimes are shown in Figure 7.3. It is easy to see that a higher correlation regime is related to higher volatilities and that correlation regimes do reoccur. Interestingly, the highest correlation regime, the fourth regime which occurs during the 1987 financial crisis, is rather short lived and spiky. This is consistent with the common idea that the correlation between financial markets increases sharply during times of economic crisis. We restricted our attention to these two different periods in separate analysis because the number of correlation regimes explodes with the increased amount of data, and the shorter intervals allow us to give a concise and illustrative discussion. Of course a more inclusive analysis would be appropriate for policy makers.

7.4 Portfolio Allocation

In Chapter 6, we proposed a re-balancing strategy where a portfolio is re-balanced whenever a regime switch is detected. Simulation studies show that on one hand, with significantly lower portfolio update frequency, our strategy achieves a Sharpe Ratio as large as the Sharpe Ratio using a daily update framework; on the other hand, compared to the regular adjustment schemes adopted in practice, our strategy produces a much higher Sharpe Ratio under similar number of adjustments. Given the positive performance using simulated data, we consider this strategy using real data.

We apply our proposed re-balancing strategy to the indices data from 1980/1/1 to 1991/07/01. We choose this period for illustration because it represents a more typical example of the financial market; the recent indices data (1993/06/26 to 2008/9/28) is not representative because each correlation regime appears only once. In addition, the mean return of Japan indices is negative for the recent period of time, which might produce negative Sharpe Ratios according to the Markowize portfolio optimization rule. In that case, we have both long and short positions of a portfolio and it does not make sense to compare the mean Sharpe Ratio.

The portfolio allocation results, which are shown in Table 7.7, are based on using different models in combination with different portfolio adjustment strategies. The portfolio re-balance strategies include daily update, regular update, and

regime-switching update. The models we compare include

- 2 CV: Regime switch in both the correlation and the volatilities, where 2 regimes are specified for the volatilities and 2 regimes are specified for the correlation;
- 3 CV: Regime switch in both the correlation and the volatilities, where 2 regimes are specified for the volatilities and 3 regimes are specified for the correlation;
- 4 CV: Regime switch in both the correlation and the volatilities, where 2 regimes are specified for the volatilities and 4 regimes are specified for the correlation;
- 4 C Switch: Regime switch in the correlation, where 4 regimes are specified for the correlation and a standard, non regime switching stochastic volatility model is used.
- 2 V Switch: Regime switch in the volatilities, where 2 regimes are specified for the volatilities and we assume a constant correlation.

The average Sharpe Ratios which are calculated based on daily re-balancing are presented in the first panel of Table 7.7 (there are about 3000 re-balance points using the daily re-balancing, which would be costly in practice). The second panel shows the results with re-balancing on a fixed interval, while the third panel presents the Sharpe Ratios under our proposed regime-switching portfolio allocation strategy. For example, in the "4 CV" model column, if we re-balance portfolios daily, the average Sharpe Ratio is 0.101. When we re-balance the portfolio whenever a switch in the volatility or the correlation structures happens, the Sharpe Ratio drops to 0.0946 with a total of 59 adjustments. For a fair comparison, we re-balance every 50 days (which leads to 60 re-balancing times, approximately the same number as in the regime switching re-balancing strategy), which results in a Sharpe Ratio, from the "4 CV" model, of 0.924. Even if we increase the re-balance frequency so that the number of adjustment increases to 75, the Sharpe ratio only increases to 0.0931, and is lower than the regime switching re-balancing

Sharpe Ratio of 0.0946. Similar relationships exist for all the other models, suggesting that the regime switching re-balancing strategy is optimal with respect to balancing Sharpe Ratio and reducing cost.

Daily Re-balancing					
	2 CV	4 C Switch	2 V Switch	3 CV	4 CV
Sharpe Ratio	0.100	0.0985	0.0987	0.100	**0.101**
Regular Re-balancing					
Re-balance Every 50 Day					
Re-balance Number:	60	60	60	60	60
Sharpe Ratio	0.0917	0.0905	0.0906	0.0915	**0.0924**
Re-balance Every 40 Day					
Re-balance Number:	75	75	75	75	75
Sharpe Ratio	0.0928	0.0917	0.0916	0.0927	**0.0931**
Regime Switching Re-balancing					
Re-balance Number:	55	21	33	57	59
SR	0.0941	0.089	0.093	0.0947	**0.0946**

Table 7.7. Portfolio allocation performance based model comparison results for international indices data during the period 1980-1991

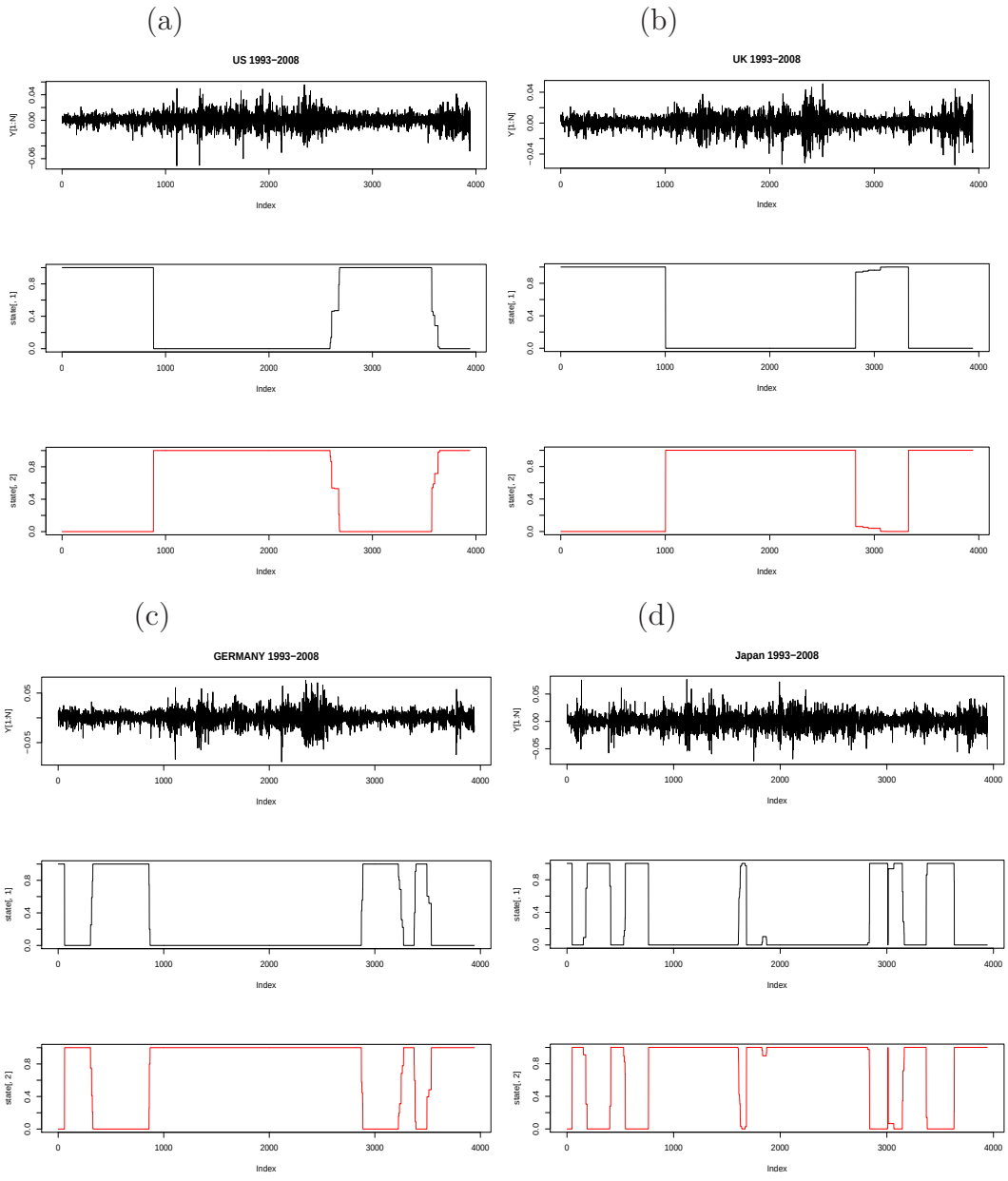

Figure 7.1. International indices log return data vs. posterior volatility regime probabilities: (a) US (b) UK (c) GERMANY (d) JAPAN

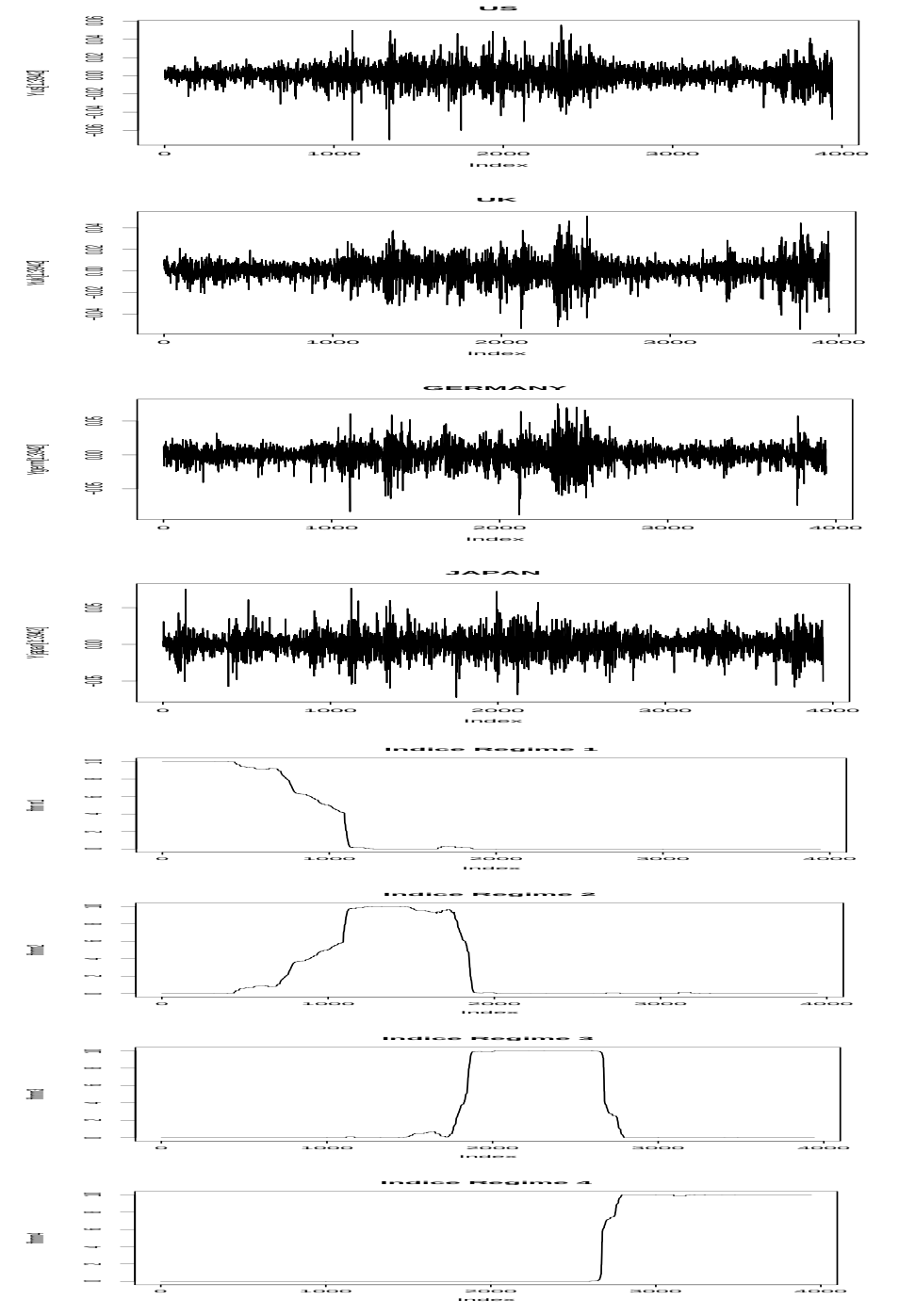

Figure 7.2. International Indices return data vs posterior probability for correlation regimes during the period of 1993-2008. The first four plots are the log return series for US, UK, GERMANY and JAPAN. The four plots below present the posterior probabilities of correlation regimes from the lowest average correlation regime to the highest.

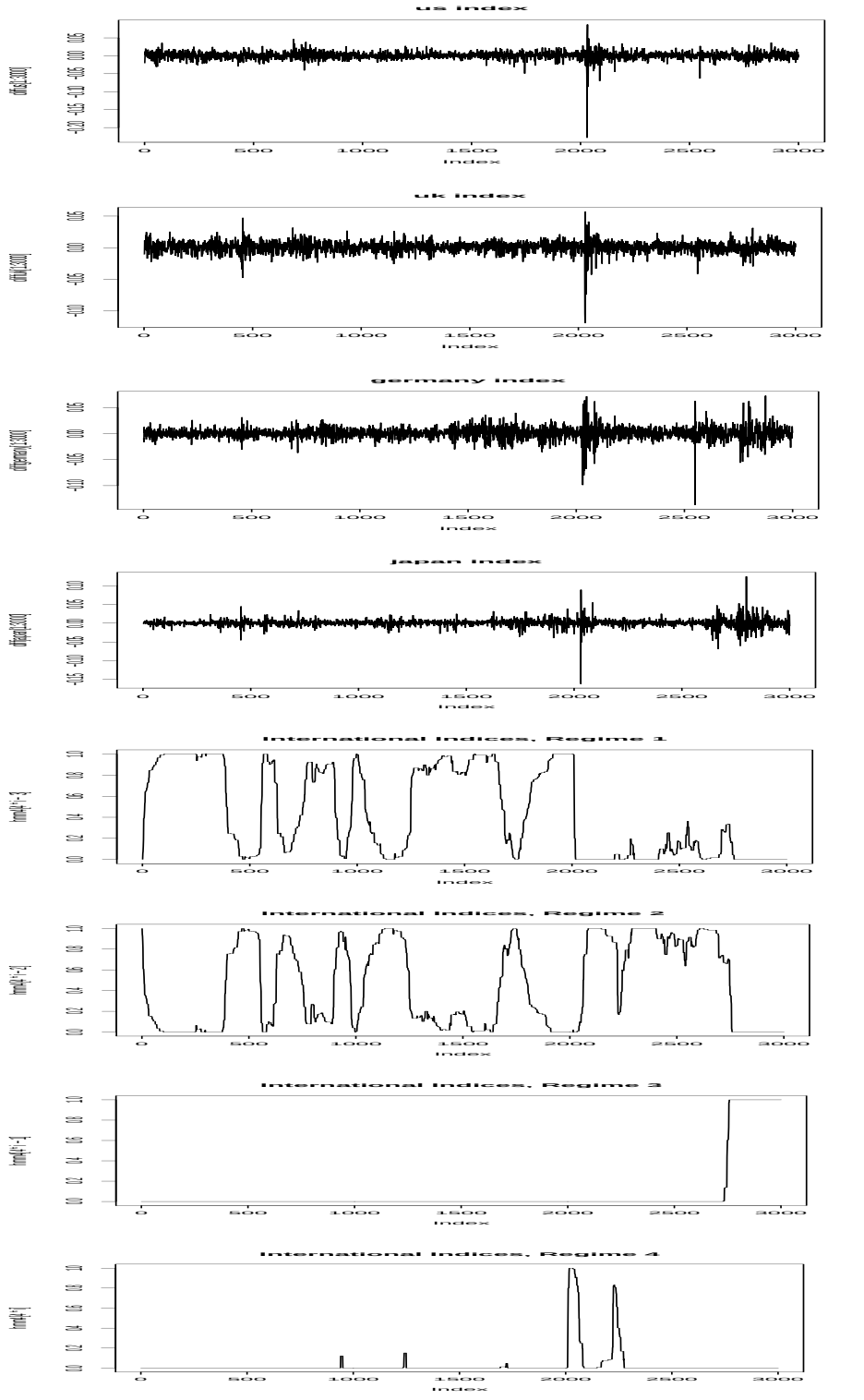

Figure 7.3. International Indices return data vs posterior probability for correlation regimes during the period of 1980-1991. The first four plots are the log return series for US, UK, GERMANY and JAPAN. The four plots below present the posterior probabilities of correlation regimes from the lowest average correlation regime to the highest. The highest correlation regime, the fourth one, corresponds to extremely high volatilities observable in log return plots. This happened in the 87' financial crisis.

Chapter 8

Conclusion and Future Work

8.1 Conclusion

In this thesis, we proposed a model and its related Bayesian estimation methodology to simultaneously model regime switching stochastic volatilities and regime switching correlations. Although this model has been mentioned in the literature, it has not been introduced and explained because it was perceived to be too complex and potentially not stable. However, we introduced an MCMC algorithm which is able to accurately estimate parameters, recover hidden regime processes, and impute missing data. Comparisons of our model and other existing models illustrate that the estimation of our model is stable and provides a better fitting of dynamic covariance matrices, which can potentially improve the performances of option pricing, risk management and portfolio allocation. The joint modeling of regime switching processes in both volatilities and correlations has other potential useful applications: looking backward, it can be used to identify structural switches associated with historic events and the relationship between volatility regimes and correlation regimes, which is helpful in explaining econometric phenomena. Looking forward, we can potentially forecast financial crisis which is characterized by the simultaneous occurrence of a high volatility and high correlation regime. In this study, we applied our model to daily indices data from four major financial markets, which offered interesting insights from an econometrics perspective. We observed that high volatility regimes are related to high correlation regimes and a switch in the volatility regimes has certain correspondence with a switch in

correlation regimes. Using this analysis, we were able to offer explanations that tied switches to historical events. We also explored another interesting application of our model, such as a proposed dynamic portfolio allocation strategy where a portfolio is re-balanced whenever a regime switch in either the correlation or volatility processes occurs. Both simulation and empirical analysis suggested that this strategy is optimal with respect to maximizing a Sharpe Ratio and reducing cost.

The general framework of this model and estimation method is as follows: first, covariances are decomposed into correlations and univariate volatilities, both of which follow regime switching processes. There is one independent regime process for each volatilities process and one for the correlation process; second, for volatilities, we extend the standard stochastic volatility model to allow for different volatility parameters, such as mean level, persistency and variance of volatilities, across different regimes; third, for correlations, the correlation matrix is assumed to be constant within a regime, but changes when each regime changes; fourth, we use a MCMC sampler where in each sweep of update, we sample univariate volatilities for each asset conditional on the remaining parameters and then estimate correlation matrix conditional on the updated volatilities. The novelty of the proposed MCMC algorithm resides in three aspects which we summarize below.

First, for univariate stochastic volatilities, we generalized the latent regime switching process $\mathbf{D}$, allowing it to be non-Markovian, which provided a better fit for data that have somewhat regular switches in regimes and hence consistent length of waiting times. We broke from traditional discrete-time Markovian modeling approaches and used a continuous-time model of $\mathbf{D}$, which is assumed to consist of jumps and waiting times. We chose continuous-time version because it is easily extended to allow $\mathbf{D}$ to be non-Markovian. In this study, we explored one of the possible extensions where the waiting times follow a gamma distribution and as a result, the regime process is non-Markovian unless the shape parameter for the gamma distribution is one. We used a reversible jump MCMC method to make inference for $\mathbf{D}$ and determined the number of regimes using several model comparison tools. We apply the non-Markovian model to natural gas future data, which is a typical example of data that have frequent regular switches in the volatilities. In this example we demonstrate that this extension is able to capture the frequent

switches in the underlying regime process and provide better fitting for the natural gas type of data with strong seasonal components.

Second, we adopted a Bayesian hierarchical model of correlation matrix which does not suffer from the technical limitations in existing correlation models and facilitated the identification of correlation regime switching process. We assumed that the off-diagonal elements of a correlation matrix shrink towards a common mean. In order to minimize computational problems associated with the PSD constraints in the correlation estimation, we introduced a layer of "shadow" priors following [35]. With this hierarchical structure, each correlation regime is uniquely identifiable and clearly represented by the average correlation hyper-parameter $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, ..., \tilde{\mu}_{K_c})$, which is an ordered vector and corresponds to the 1^{st} to the K_c^{th} correlation regimes, indicating the lowest to the highest average correlation among a set of assets. In addition, using our method, we do not have to make any unrealistic assumptions for the correlation matrix, e.g., restricting all correlation elements to be positive as in [44]. This is especially important for estimating correlations at extreme times, e.g., financial crisis, where assets are clearly negatively correlated.

Third, we proposed a Bayesian imputation method which extracts information from the estimated volatility and correlation structures and makes inferences for missing values in a multivariate time series. This method is easy to implement and can provide more accurate parameter estimations for missing data when compared with other popular imputation methods used in practice.

In addition to the three points above, our model has great flexibility. The number of regimes K can take any value, and is pre-specified for either the volatility or the correlation switching processes before we implementing the full MCMC algorithm. In addition, all the volatility parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\tau}^2)$ are allowed to differ by regimes. This flexibility provides us a platform to find the most suitable structure to describe a data set and explore interesting econometric phenomena. For example, we compared different correlation regime numbers and concluded that a four correlation regime model fit the recent international indices data the best. However, if we do not have the freedom to choose K, we might have ended in a wrong conclusion of only two correlation regimes. The failure in identifying the third and fourth correlation structures might mislead financial practitioners to use

wrong parameters for calculating option prices, or risk metrics, which would be costly. Another example is that the flexibility to identify whether $\boldsymbol{\phi}$ and $\boldsymbol{\tau}^2$ differ across regimes helped us to realize the fact that for indices data, a high volatility regime is related to a higher persistence and lower variance of the volatility.

With a variety of competing models at hand for each data set, we have the question of how to determine the best model. The model comparison tools we used in this study include: marginal likelihood, DIC value, and posterior model probability. Because of the unobserved regime processes $\mathbf{D}$ and volatility processes $\mathbf{h}^i$, the evaluation of marginal likelihood is done with Hamilton filter or by calculating the harmonic mean of conditional likelihoods. The DIC comparison criterion which penalizes complex models is implemented because it is computationally simple within our MCMC estimation framework. We also adopt a reversible jump MCMC method which explores and compares models with different parameter spaces and provides a posterior probability for each model. The model with the highest probability is the best fitting model. In our simulation studies, all the model comparison tools reach the same conclusion and are capable of identifying the true model. In some applications on real data, the DIC value points to a simpler model which is not favored by other likelihood based methods. Nevertheless, the choice of the best fitting model when discrepancy occurs among different comparison tools depends on our investigation objectives: we would choose the model with the largest DIC if we are very sensitive to and want to avoid too complex models, while the model with the largest marginal likelihood is preferred if our purpose is to find the best fitting model with some tolerance of complexity.

An appealing application of our model is with respect to dynamic portfolio allocation. A portfolio manager determines the proportion of each asset in a portfolio with the goal of increasing return and reducing risk. In practice a portfolio manager adjusts or re-balances the asset weights based on recent data and information. However, we observed that it is a waste of money, with respect to transaction cost to re-balance a portfolio when there is not much change in the structure of asset values. To address the problem, we proposed a strategy which adjusts portfolio weights whenever a regime switch occurs in either volatilities or correlations. The motivation of our proposal is straightforward but appears to have never been implemented in practice. A possible explanation is that no previous work can jointly

model regime switches in both volatilities and correlations. Both simulation and empirical analysis show that our strategy saves re-balance cost and increases average Sharpe Ratios compared to regularly timed portfolio re-balancings.

In conclusion, we provided a framework for representing and estimating dynamic covariance matrices, where regime switching processes for volatilities and correlations are modeled simultaneously. We find important application in econometric analysis and portfolio allocation and we also anticipate that these models will lead to interesting applications in risk management and asset pricing in the future.

8.2 Future Work

Our dynamic covariance model, along with the Bayesian hierarchical correlation estimation method and the continuous-time modeling of regime switching process, leaves plenty of room for extension and addressing practical problems. We list several topics of interest in this section.

8.2.1 Leverage Effect-Perfect Storm

Leverage effect refers to the financial phenomenon that a decrease in asset price tends to lead to an increase in volatility. When this asymmetric feature is not permitted in a stochastic volatility model, option prices could be substantially biased [26]. Leverage effect can be included into our stochastic volatility model simply by specifying a negative correlation between the two shock terms in Equation 3.1, that is $corr(\varepsilon_t, \eta_t) = \rho_t$ where $\rho_t < 0$.

Although easily incorporated into the model, the estimation of leverage effect is not straightforward. Recall that we estimate $\mathbf{h}$ in our univariate regime switching stochastic volatility model by approximating the log Chi-square distributed error term in Equation 3.14 with a mixture of normal distributions. However, when the leverage effect is considered, such an approximation is not valid. More sophisticated method need to be used [50].

The possibility of modeling regime switching in volatilities, correlations and leverage effects simultaneously motivates us to think of a "Perfect Storm" model.

The terminology "Perfect Storm" is introduced to describe an event where a rare combination of circumstances aggravate a situation drastically [1]. In the financial context, we link a "Perfect Storm" to financial crisis which is risky and potentially destructive. A more concrete definition is a market condition where some assets in a portfolio have a sharp drop in prices, together with an increased leverage effect, which gives rise to very large, but upward volatilities. The correlations among assets in this portfolio increase at the same time, which further drags down the prices of all the other assets. The 2008 subprime crisis is a salient example of our "Perfect Storm". Consequently an ability to identify and potentially predict such a market state would be invaluable to the market, investors and regulators. Therefore, in the future it would be interesting to incorporate leverage effects into our model and try to forecast a "Perfect Storm".

8.2.2 Group correlation

As mentioned in [35], the shrinkage analysis of correlation matrix not only introduces a common mean for all the correlation elements, but also allows grouping of variables according to natural clusters of underlying variables or similarities among the correlations. The grouping improves model fitting because prior information can be incorporated into estimation. Moreover, the assumption of strict dependence on the same prior in standard models can be mitigated by grouping.

Assume there are Q groups of variables. A grouped correlation model contains group information for each element in the correlation matrix [35], which is shown below:

$$f(\Gamma|\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\sigma}}^2, \boldsymbol{\vartheta}) = C(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\sigma}}^2, \boldsymbol{\vartheta}) \prod_{i<j} [\sum_{q=1}^{Q} I(\vartheta_{ij} = q) \exp\{-(\rho_{ij} - \tilde{\mu}_q)^2/(2\tilde{\sigma}_q^2)\}] I(\Gamma \in R^p) \tag{8.1}$$

where $\boldsymbol{\vartheta}$ is a $p * p$ matrix indicating the group information. $I(\vartheta_{ij} = q)$ indicates that the (i, j)th element in the correlation matrix belongs to group q. $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, ..., \tilde{\mu}_Q), \tilde{\boldsymbol{\sigma}}^2 = (\tilde{\sigma}_1^2, ..., \tilde{\sigma}_Q^2)$ are the hyper prior parameters for each group.

A slightly different specification is called grouped variable model, which clusters

[1] Wikipedia

assets instead of correlation elements:

$$f(\Gamma|\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\sigma}}^2, \boldsymbol{\vartheta}) =$$
$$C(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\sigma}}^2, \boldsymbol{\vartheta}) \prod_{i<j} \left[\sum_{q_1 q_2} I(\vartheta_i = q_1) I(\vartheta_j = q_2) \exp\{-(\rho_{ij} - \tilde{\mu}_{q_1 q_2})^2 / (2\tilde{\sigma}^2_{q_1 q_2})\} \right] I(\Gamma \in R^p) \quad (8.2)$$

A combination of regime switching and grouping in the correlation modeling can result in a wide range of interesting applications. For example, the 2008' financial crisis leads to a fundamental change in the financial market. Not only did assets become more correlated, but also the grouping indicators could have changed. Due to nationalization, several major investment banks in the US, Goldman Sachs for example, are possibly grouped with major commercial banks now. Therefore, a regime switch in the correlation could result in a change in the correlation values, or a switch in the group information, or both. Such changes would impact the decision of portfolio managers. In this example, we have evidence that the grouping information should have changed because of the new policy. However in most other cases, we do not have such a clear message in determining a possible switch in groups. This is especially true for companies which are expanding or changing their services. Those companies need a model to decide whether their transformation is successful, when they achieve their transformation goal, etc. The Enron example analyzed in a static correlation framework in [35] could be re-analyzed by our dynamic models. Instead of checking whether Enron is an energy company or a financial company, we can try to answer whether Enron, before its demise, successfully transformed to be a financial company from an energy company, and approximately when such a change happened.

8.2.3 Relationship of Multiple Hidden Regime Processes

In the thesis, we model multiple regime switching processes for volatilities and correlations simultaneously. However, we make an implicit assumption that these processes are independent among each other: a switch in one chain neither receive from nor extend influence to other chains. However, empirical evidences suggest the existence of complex relationships among the regime switching processes. For example, several phenomena have been well documented in econometrics literature:

* Spillover: A switch in regime of a dominating asset leads to a switch in the dominated asset (with a lag);

* Interdependence: A reciprocal spillover;

* Comovement: A contemporaneous change in regimes.

A model which can explain such relationships should be able to improve model fitting and forecasting performance in the future. A possible solution to relate the hidden regime processes is to relate the waiting time parameters. Following the notations in Chapter 5, we assume there are p financial assets in a portfolio and K possible regimes for the volatilities. The volatility regime switching processes are denoted as $\mathbf{D}^{hi}$, $i = 1, ..., p$. The correlation regime switching process for this portfolio is $\mathbf{D}^{c}$. The waiting time parameters for the i^{th} asset in the k^{th} regime is denoted by $(\alpha_k^{hi}, \beta_k^{hi})$. The waiting time parameters for $\mathbf{D}^{c}$ in the k^{th} regime is denoted by $(\alpha_k^{c}, \beta_k^{c})$. We list possible extensions to our model case by case.

Case 1: Assume $\alpha_k^{h1} = ... = \alpha_k^{hp}$, $\beta_k^{h1} = ... = \beta_k^{hp}$ $k = 1, ..., K$. That is, the waiting times are generated from the same distribution with the same parameters. In this case, each $\mathbf{D}^{hi}$ is a different realization from the same distribution.

Case 2: Assume $\alpha_k^{hi} \sim N(\mu^{\alpha}, \sigma^{\alpha})$, $\beta_k^{hi} \sim N(\mu^{\beta}, \sigma^{\beta})$ $k = 1, ..., K$. That is, for each regime, the waiting time parameters shrink towards the same hyper prior distributions.

Case 3: Empirically, when the volatilities are in a high state, the correlations increase as well. Therefore, $(\alpha_k^{c}, \beta_k^{c})$ should be related to $(\alpha_k^{hi}, \beta_k^{hi})$ in certain ways and we could explore that in more detail.

Case 4: The waiting time parameters $(\boldsymbol{\alpha}^{hi}, \boldsymbol{\beta}^{hi})$, and the transition matrix $\mathbf{P}^{hi}$ can be modeled as a function of covariates. For example, we can model the transition probability for asset 1 by $p(D_{t+1}^{h1}|D_t^{h1}) = a_1 + \sum_{l=2}^{p} a_l I(D_t^{hl} \neq D_{t-1}^{hl})$, where $I(D_t^{hl} \neq D_{t-1}^{hl})$ is an indicator function telling whether there is a switch in volatility regime at previous time points of other assets. In this way, we relate multiple regime switching processes so that they influence each other.

All the four cases mentioned above point towards a general modeling and inference problem with respect to multiple hidden regime switching processes, or more general, with respect to multivariate point processes. In essence, the statistical question reduces to determining how to "shrink" or effectively reduce the dimension of these multivariate processes. For example, while there may well be one

hidden regime switching process which is driving the correlation structure between a set of assets and p hidden processes which are driving the individual variance (stochastic volatility) of each asset, what are the appropriate relationships between these processes. While we have not found extensive literature treating this problem of multivariate point processes, we view this as a fundamentally important problem.

8.2.4 Regime Switching Applicable to Other Models

We proposed to allow for latent non-Markovian regime switching processes for covariances in our study. However, the non-Markovian structure should have a much more general application in time series analysis. For example, there are a variety of time series data which have evident seasonal effects in the means (instead of variances),e.g., weather data, energy data. Our proposed non-Markovian regime switching model should be able to improve upon existing Markovian modeling for them. Moreover, models such as regime switching CAPM model, [25] [9] and regime switching factor models [38] are based on the Markovian switching assumption, which could be relaxed in our modeling framework as well.

In addition, we are very interested in ensuring that any models that we develop are able to help answer practical problems encountered in the Finance industry. Hence, we will hold up any modeling advances to the criteria of improving practical understanding.

Appendix A

Signal Simulation Smoother

The signal simulation smoother consists of two steps: forward filtering and backward sampling (FFBS).

A.1 Forward filtering

A Kalman filter is an efficient recursive filter that estimates the state variables of a state space model from a series of noisy measurements [1].

A state space model is like this:

$$\begin{aligned} m_t &= H_t x_t + v_t \\ x_t &= F_t x_{t-1} + B_t u_{t-1} + w_{t-1} \end{aligned} .$$

where x_t is the state variable, v_t and w_t are independent normal variables and $v_t \sim N(0, R_t)$, $w_t \sim N(0, Q_t)$. In the context of our model, we have

$$\begin{aligned} y_t^* &= h_t + z_t \\ h_t &= \phi_{D_t} h_{t-1} + \mu_{D_t}(1 - \phi_{D_t}) + \tau_{D_t}\eta_t \end{aligned} .$$

We can easily identify the correspondence between our model and the classic state space model, where $m_t = \log(y_t^2) = y_t^*$, $H_t = 1$, $x_t = h_t$, $F_t = \phi_{D_t}$, $B_t u_{t-1} = \mu_{D_t}(1 - \phi_{D_t})$ and $Q_t = \tau_{D_t}^2$. In addition, conditional on the indicator variable s_t, R_t equals one of the ν_i^2 listed in Table 3.1.

[1] http://en.wikipedia.org/wiki/Kalman_filter

A Kalman filter consists of two phases: a "Predict" phase and an "Update" phase. The predict phase uses previous estimation result to produce an estimator for current time point. Then in the update phase, measurement information for current time point is used to refine the prediction. To be more specific,

Predict Phase:

$$\begin{aligned}\hat{x}_{t|t-1} &= F_t\hat{x}_{t-1|t-1} + B_t u_{t-1} \\ P_{t|t-1} &= F_t P_{t-1|t-1} F_t^T + Q_{t-1}\end{aligned} \quad .$$

where $\hat{x}_{t|t-1}$ is the predicted value for current time, and $\hat{x}_{t-1|t-1}$ is the updated value from last iteration. $P_{t|t-1}$ is the predicted estimation covariance.

Update Phase:

$$\begin{aligned}\tilde{y}_t &= m_t - H_t\hat{x}_{t|t-1} \\ S_t &= H_t P_{t|t-1} H_t^T + R_t \\ K_t &= P_{t|t-1} H_t^T S_t^{-1} \\ \hat{x}_{t|t} &= \hat{x}_{t|t-1} + K_t\tilde{y}_t \\ P_{t|t} &= (I - K_t H_t) P_{t|t-1}\end{aligned} \quad .$$

where $\tilde{y}_t$ is the measurement residual, S_t is the residual covariance and K_t is the Kalman gain. $\hat{x}_{t|t}$ is the updated value for the current iteration, and $P_{t|t}$ is the updated estimation covariance. The key element in this step is the Kalman gain K_t, which is a kind of weight chosen to minimize $P_{t|t}$, the updated estimation covariance.

A.2 Backward Sampling

Backward sampling is a smoothing procedure to extract the hidden state variable information conditional on the complete series of measurements $\mathbf{m} = (m_1, ...m_n)$. In the forward sampling pass, residuals $\tilde{y}_t$, covariance for residuals S_t, and the Kalman gain K_t are stored. Setting $r_n = 0, N_n = 0$, and writing $D_t = S_t^{-1} +$

$K_t N_t K_t$ and $n_t = S_t^{-1}\tilde{y}_t - K_t r_t$, we run backward from $t = n, n-1, ..., 1$.

$$\begin{aligned} C_t &= R_t - R_t D_t R_t, \\ \kappa_t &\sim N(0, C_t), \\ r_{t-1} &= H_t S_t^{-1}\tilde{y}_t + L_t' r_t - V_t' C_t^{-1}\kappa_t \qquad . \\ V_t &= R_t(D_t H_t - K_t' N_t B_t), \\ N_{t-1} &= H_t' S_t^{-1} H_t + L_t' N_t L_t + V_t' C_t^{-1} V_t \end{aligned}$$

where $L_t = B_t - K_t H_t$. Then $m_t - R_t n_t - \kappa_t$ is a draw from the multivariate normal posterior distribution $p(H_t x_t | \mathbf{m}, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes all the model parameters.

Appendix B

Augmented Auxiliary Particle Filter

The objective of an augmented auxiliary particle filter is to calculate the marginal log-likelihood $\log f(\mathbf{y})$ which can be expressed successively by

$$\log f(\mathbf{y}) = \log(y_1) + \sum_{t=2}^{N} \log f(y_t|\mathbf{y}_{t-1}) \tag{B.1}$$

where $\mathbf{y}_{t-1} = (y_1, ..., y_{t-1})$.

$$f(y_t|\mathbf{y}_{t-1}) = \int f(y_t|h_t) f(h_t|\mathbf{y}_{t-1}) dh_t \tag{B.2}$$

and

$$f(h_t|\mathbf{y}_{t-1}) = \int f(h_t|h_{t-1}) f(h_{t-1}|\mathbf{y}_{t-1}) dh_{t-1} \tag{B.3}$$

Therefore, in our model, given a sample $h_{t-1}^{(i)}, i = 1, ..., M$ of size M from $f(h_{t-1}|\mathbf{y}_{t-1})$, $h_t^{(i)}, i = 1, ..., M$, can be sampled from

$$h_t^{(i)}|h_{t-1}^{(i)} \sim N(\mu_{D_t} + \phi_{D_t}(h_{t-1}^{(i)} - \mu_{D_t}), \tau_{D_t}^2) \tag{B.4}$$

and $f(y_t|\mathbf{y}_{t-1})$ can be approximated by $\frac{1}{M}\sum_{i=1}^{M} f(y_t|h_t^{(i)})$, where $h_t^{(i)}|h_{t-1}^{(i)}$ are named filtered particles.

The question is how to sample from $f(h_t|\mathbf{y}_t)$. We can do it iteratively: by the

Bayes rule and Equation B.3, we have

$$\begin{aligned} f(h_t|\mathbf{y}_t) &\propto f(y_t|h_t)f(h_t|\mathbf{y}_{t-1}) \\ &\propto f(y_t|h_t)\frac{1}{M}\sum_{i=1}^{M} f(h_t|h_{t-1}^{(i)}) \\ &\propto \frac{1}{\sqrt{\exp(h_t)}}\exp\frac{-y_t^2}{2\exp(h_t)}\frac{1}{M}\sum_{i=1}^{M}\exp\frac{-(h_t-\mu-\phi(h_{t-1}^{(i)}-\mu))^2}{2\tau^2} \end{aligned} \tag{B.5}$$

Then we can sample h_t from this distribution either by an importance sampling or an accept-reject sampling method. In our experience, $M = 3000$ is good enough for an approximation for $\log f(\mathbf{y})$.

Bibliography

[1] M. Asaia and M. McAleer. The structure of dynamic correlations in multivariate stochastic volatility models. *Journal of Econometrics*, 150(2):182–192, 2009.

[2] O. P. Attanasio. Risk, time-varying second moments and market efficiency. *Review of Economic Studies*, 58(3):479–94, May 1991.

[3] X. Bai, J. R. Russell, and G. C. Tiao. Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *Journal of Econometrics*, 114(2):349–360, June 2003.

[4] L. Bauwens, C. Bos, and H. van Dijk. Adaptive polar sampling with an application to a bayes measure of Value-at-Risk. *Tinbergen Institute, Discussion Paper TI 99-082/4*, 1999.

[5] A. Berg, R. Meyer, and J. Yu. Deviance information criterion for comparing stochastic volatility models. *Journal of Business and Economic Statistics*, 22(1):107–20, January 2004.

[6] M. Billio and M. Caporin. Multivariate markov switching dynamic conditional correlation GARCH representations for contagion analysis. *Statistical Methods and Applications*, 14(2):145–161, 11 2005.

[7] T. Bollerslev. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69(3):542–47, August 1987.

[8] T. Bollerslev. Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *The Review of Economics and Statistics*, 72(3):498–505, August 1990.

[9] L. Cappiello and T. A. Fearnley. International CAPM with regime switching GARCH parameters. FAME Research Paper Series rp17, International Center for Financial Asset Management and Engineering, July 2000.

[10] C. K. Carter and R. Kohn. On Gibbs sampling for state space models. *Biometrica*, 81(3):541–553, 1994.

[11] C. M. Carvalho and H. F. Lopes. Simulation-based sequential analysis of markov switching stochastic volatility models. *Computational Statistics and Data Analysis*, 51(9):4526–4542, May 2007.

[12] F. Chesnay and E. Jondeau. Does correlation between stock returns really increase during turbulent periods? *Economic Notes*, 30:53–80, 2003.

[13] S. Chib. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90:1313–1321, 1995.

[14] S. Chib, Y. Omori, and M. Asai. Multivariate stochastic volatility. CIRJE F-Series CIRJE-F-488, CIRJE, Faculty of Economics, University of Tokyo, 2007.

[15] P. De Jong and N. Shephard. The simulation smoother for time series models. *Biometrika*, 82(2):339–350, 1995.

[16] R. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. In *Econometrica*, volume 50, pages 987–1008, 1982.

[17] R. Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, 20(3):339–50, July 2002.

[18] R. F. Engle and G. Gonzalez-Rivera. Semiparametric ARCH models. *Journal of Business and Economic Statistics*, 9:345–360, 1991.

[19] J. Geweke. Bayesian analysis of stochastic volatility models: Comment. *Journal of Business and Economic Statistics*, 12(4):397–99, October 1994.

[20] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in practice*. Chapman & Hall, 1996.

[21] P. J. Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82:711–732, 1995.

[22] J. D. Hamilton and R. Susmel. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64(1-2):307–333, 1-2 1994.

[23] P. Hansen and A. Lunde. Consistent preordering with an estimated criterion function, with an application to the evaluation and comparison of volatility models. Working papers, Brown University, Department of Economics, 1 2003.

[24] A. Harvey, E. Ruiz, and N. Shephard. Multivariate stochastic variance models. *Review of Economic Studies*, 61(2):247–64, April 1994.

[25] H.-C. Huang. Tests of regime-switching CAPM. *Applied Financial Economics*, 10(5):573–78, October 2000.

[26] J. C. Hull and A. D. White. The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42(2):281–300, June 1987.

[27] S. Hwang, S. E. Satchell, and P. L. V. Pereira. How persistent is volatility? an answer with stochastic volatility models with markov regime switching state equations. *Journal of Business Finance and Accounting*, 34(5-6):1002–1024, June/July 2007.

[28] E. Jacquier, N. G. Polson, and P. E. Rossi. Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 12(4):371–89, October 1994.

[29] E. Jacquier, N. G. Polson, and P. E. Rossi. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, 122(1):185–212, September 2004.

[30] N. L. Johnson. An approximation to the multinomial distribution some properties and applications. *Biometrika*, 47(1-2):93–102, 1960.

[31] C.-J. Kim. Dynamic linear models with markov-switching. *Journal of Econometrics*, 60(1-2):1–22, 1994.

[32] S. Kim, N. Shephard, and S. Chib. Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65(3):361–93, July 1998.

[33] C. G. Lamoureux and W. D. Lastrapes. Persistence in variance, structural change, and the GARCH model. *Journal of Business and Economic Statistics*, 8(2):225–34, April 1990.

[34] S. Laurent, L. Bauwens, and J. V. K. Rombouts. Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1):79–109, 2006.

[35] J. C. Liechty, M. W. Liechty, and P. Muller. Bayesian correlation estimation. *Biometrika*, 91(1):1–14, March 2004.

[36] J. C. Liechty and G. O. Roberts. Markov chain monte carlo methods for switching diffusion models. *Biometrika*, 88:299–315, 2001.

[37] H. F. Lopes and M. West. Bayesian model assessment in factor analysis. *Statistica Sinica*, pages 41–67, 2004.

[38] H. Lopesa and C. Carvalho. Factor stochastic volatility with time varying loadings and markov switching regimes. *Journal of Statistical Planning and Inference*, 137(10):3082–3091, 2007.

[39] B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 1963.

[40] H. M. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

[41] D. B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370, 1991.

[42] M. Newton and A. Raftery. Approximate bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society, Series B*, 56(1):3–48, 1994.

[43] E. N. Ole E. Barndorff-Nielsen and N. Shephard. Some recent developments in stochastic volatility modelling. *Quantitative Finance*, 2(1):11–23, Feburary 2002.

[44] D. Pelletier. Regime switching for dynamic correlations. *Journal of Econometrics*, 131(1-2):445–473, 2006.

[45] M. K. Pitt and N. Shephard. Filtering via simulation: auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599, 1999.

[46] S.-H. Poon. Extreme value dependence in financial markets: Diagnostics, models, and financial implications. *Review of Financial Studies*, 17(2):581–610, 2004.

[47] N. Shephard. *Stochastic Volatility*. Oxford University, 2005.

[48] D. R. Smith. Markov-switching and stochastic volatility diffusion models of short-term interest rates. *Journal of Business and Economic Statistics*, 20(2):183–97, April 2002.

[49] M. So, K. Lam, and W. K. Li. A stochastic volatility model with markov switching. *Journal of Business and Economic Statistics*, 16(2):244–253, 4 1998.

[50] J. Yu. On leverage in a stochastic volatility model. *Journal of Econometrics*, 127(2):165–178, August 2005.

[51] J. Yu and R. Meyer. Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews*, 25(2-3):361–384, 2006.

Vita

Lu Zhang

RESEARCH INTERESTS

Bayesian analysis, Monte Carlo Markov Chain (MCMC), Dynamic linear model, Hidden Markov model, Time series data analysis, Data mining

EDUCATION

- The Pennsylvania State University, University Park, PA — May 2010
  Ph.D Candidate of Statistics — GPA: 3.95/4.0
- The Pennsylvania State University, University Park, PA — Dec 2008
  M.S. of Statistics — GPA: 3.95/4.0
- Chu Kochen Honors College, Zhejiang University, Zhejiang, China — Jul 2005
  B.S. of Statistics — GPA :3.93/4.0

PUBLICATIONS

- Automatic Tag Recommendation Algorithms for Social Recommender Systems (Transactions on the Web (TWEB) 2009)
- A Non-parametric Approach to Pair-wise Dynamic Topic Correlation Detection (ICDM 2008)
- A Sparse Gaussian Processes Classification Framework for Fast Tag Suggestions (CIKM 2008)
- Master Thesis: Regime Switching Stochastic Volatility and its Empirical Analysis.

INTERNSHIP

HSBC Bank USA, Buffalo, NY, Summer 2008

- Time series forecasting models for Online Saving balances
- Logistic regression and classification models for mortgage customer analysis
- Profit and loss computation and analysis for Deposit balances

PROFESSIONAL ACTIVITIES

- Poster: 2009 Joint Statistical Meetings (JSM), Washington DC
- 2009 ENAR Spring Meeting, San Antonio, Texas
- 2009 NISS Workshop in Exploring Statistical Issues in Financial Risk Modeling and Banking Regulation,DC
- Presentation: IEEE International Conference on Data Mining series (ICDM 2008),Pisa, Italy
- Presentation: 2008 Joint Statistical Meetings (JSM), Denver, Colorado
- Seminar on Bayesian Inference in Econometrics and Statistics (SBIES), Chicago, Illinois

TEACHING EXPERIENCES

- Instructor of Stat 250: Biostatistics, Spring 2008
- Instructor of Stat 200: Elementary Statistics, Summer 2007