The Pennsylvania State University
The Graduate School

REGIME SWITCHING STOCHASTIC VOLATILITY AND ITS EMPIRICAL ANALYSIS

A Thesis in
Statistics
by
Lu Zhang

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

December 2008

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

John C. Liechty
Professor of Marketing and Statistics
Thesis Advisor

Bruce G. Lindsay
Professor of Statistics
Statistics Department Head

Murali Haran
Professor of Statistics

*Signatures are on file in the Graduate School.

Abstract

We consider stochastic volatility model with regime switching. We model the unobserved regimes as a continuous-time latent or hidden Markov chain, with exponential waiting times. This allows us to identify the hidden regimes and compare the posterior probability of changes in regimes with news events. Our model is very flexible: the number of possible regimes, the mean reverting, persistence and variance parameters in the volatility equation can all be controlled by different hidden Markov chains. We use Monte Carlo Markov Chain (MCMC) method for inference; model selection is based on calculating the posterior probability of the model, given the data via a reversible jump approach. We also consider maximum likelihood, and out of sample predictive validation. We report the analysis of several empirical dataset: $S\&P$ 500, Federal rate, exchange rate, and demonstrate proposed models out perform benchmark models with regards to fit and predictive ability.

Table of Contents

List of Figures

List of Tables

Acknowledgments

This thesis arose out of one year of research with my thesis adviosr, Dr. John Liechty. I decided to enter the realm of financial statistics in Feb, 2007. This is a completely new field for me and it is exciting. I started to pick up basic knowledge in it, read through extensive literature, and with the help of my advisor, I finally decided on this topic in september, 2007. I gratefully acknowledge my advisor Dr. John Liechty for his guidance, advice, supervision and crucial contribution.

It is also my pleasure to convey my gratitude to my committee member, Dr Tim Simin, for his contribution to this thesis, and suggestions and kind discussions.

I will also gratefully thank my department, statistics department in Penn State University, where I received high level education, and found my interests in quantitative analysis. I have taken courses with Dr.Bruce Lindsay, Dr. Naomi Altman, Dr. Tom Hettmansperger, Dr. Naomi Altman, 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 is my temporary advisor for my first year graduate life and helped me to get accustomed to graduate life. And also I am indebted to Dr. Naomi Altman who I worked with in research in the summer of 2006, and she is really nice. Although I did not take course with Dr. Runze Li, I would also like to say thanks to him for his advices and guidance whenever I need help. I have also benefited by discussion and suggestions from Dr. Zhibiao Zhao. Finally and most importantly, I would thank Dr. Donald Richards and Dr. Jia Li to recruit me as Phd candidate for this wonderful department!

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

I am extraordinarily fortunate to meet my fiance, Yang Song, whose dedication, care, love and persistent confidence in me, has taken the load off my shoulder. I would also thank his parents, who have given me such an excellent lifemate.

Words fail me to express my appreciation to my parents, my father, Jiansheng Zhang, my mother Sujun Zeng, who raised me and supported me throughout for all my 25 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 fiance, Yang Song.

Chapter 1

Univariate Stochastic Volatility - Literature Review

1.1 Autoregressive Heteroscedastic (ARCH) Model

Time varying volatilities of financial asset returns are well established. In financial return time series the autocorrelation of the return variable tends to decay relatively fast while the autocorrelation of its second moment persists. Mandelbrot (1963)documented volatility clustering phenomenon, where 'large changes tend to be followed by large changes-of either sign-and small changes by small changes'.

In 1982, Engle[4] introduced one of the most popular methods, the ARCH/GARCH models, to explain the volatility characteristics. ARCH/GARCH models are discrete time models, with their main application in forecasting. ARCH/GARCH models are structured such that one-step ahead forecast is readily available and one can carry out n-step ahead prediction iteratively. Value-at-Risk (VaR)of a portfolio can be computed based on the predicted volatilities and certain assumed shock distributions. These models also have application in option pricing problems. In the following sections, We will introduce several models that are related to my research.

1.2 Basic ARCH (GARCH) model

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

$$Y_t = X_t\beta + \sqrt{h_t}Z_t; \tag{1.1}$$

$$h_t = a_0 + \sum_{i=1}^{p} a_i Z_{t-i}^2 h_{t-i}. \tag{1.2}$$

where Y_t denotes the difference of log-return. $X_t\beta$ is the mean for Y_t and Z_t is assumed to be i.i.d Normal(0,1). Note h_t is determined by the past shocks (or mean corrected log returns), a_0 is greater then 0 and a_i are assumed to be nonnegative because h_t is a variance.

Empirical experience calls for large p in the conditional variances formula because h_t has high persistency. This leads to a generalization of the ARCH model, namely GARCH(p,q), which was proposed by Bollerslev(1986)[1]. Hansen and Lunde (2003)[7], compared 330 candidate models applied to daily frequency return data with realized volatilities, and concluded that no model can outperform GARCH(1,1) when fitted to exchange rate data.

The GARCH(1,1) model is different from ARCH model in terms of the determination of the volatility term h_t.

$$h_t = a_0 + a_1 Z_{t-1}^2 h_{t-1} + b_1 h_{t-1}. \tag{1.3}$$

The GARCH (1,1) model is analogous to ARMA(1,1) model with a quadratic error term.

The fact that the distribution of financial data exhibits fat tails is of major concern with respect to the validity of ARCH/GARCH models. Even the errors are modeled using a t-distribution, such problem still remains. (Poon 2004)[18].

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

$$h_t = a_0 + \sum_{i=1}^{\infty} \phi_i Z_{t-i}^2 h_{t-i}. \tag{1.4}$$

1.3 Generalization of ARCH/GARCH model

1.3.1 ARCH model considering Leverage Effect

The leverage effect is a common phenomena in financial returns. The usual claim is that a decrease in the price (return) will increase the future expected volatility and vice versa. In basic GARCH model, because conditional volatility cannot be negative, h_t is modeled as a linear combination of squared terms (with nonnegative weights), which blinds the effect of sign of Z_t. Therefore this symmetry feature can not explain the asymmetric leverage effect (Nelson 1991)[16]

A generalization of ARCH/GARCH model to capture the leverage effects is called EGARCH model. In this model, $Ln(h_t)$ 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 volatility is:

$$Ln(h_t) = \alpha_0 + \sum_{i=1}^{q} \beta_j Ln(h_{t-i}) + \sum_{i=1}^{p} g(Z_{t-i}). \tag{1.5}$$

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

$$g(Z_t) = \theta Z_t + \gamma[|Z_t| - E(Z_t)]. \tag{1.6}$$

1.3.2 ARCH model with Switching Regions

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 such swift shift in market structures and they tend to perform poorly with respect to out of sample prediction. Basic ARCH/GARCH model implies a high degree of persistence that may not exist in empirical data. Hamilton (1994)[6] proposed a switching ARCH model (SWARCH), which explains possible structural changes in the ARCH process.

Let $\{D_t\}$ be some unobserved sequences whose variable D_t can take on the values of 1,2,...,K. They represent the underlying volatility structure in financial market. $\{D_t\}$ is assumed to be a discrete time Markov Chain with a transition matrix P_{K*K}.

A simple SWARCH model can be written for residuals u_t:

$$u_t = \sqrt{g_{D_t}} * \widetilde{u}_t \tag{1.7}$$

Hence $\widetilde{u_t}$ is assumed to follow a ARCH-L(q)process.

$$\widetilde{y_t} = \sqrt{h_t}\nu_t \tag{1.8}$$

ν_t is i.i.d unit variance sequence, and

$$h_t = a_0 + a_1\widetilde{u_{t-1}}^2 + a_2\widetilde{u_{t-2}}^2 + ... + a_q\widetilde{u_{t-q}}^2 + \xi I_{t-1}\widetilde{u_{t-1}}^2 \tag{1.9}$$

where $I_{t-1} = 1$ if $u_{t-1} \leq 0$, and $I_{t-1} = 0$ if $u_{t-1} > 0$.

This model is called a K state qth order Markov switching ARCH process, and is denoted by SWARCH (K,q), or by ARCH-L, SWARCH-L(K,q).

1.4 Stochastic Volatility Model Literature

An alternative to the ARCH framework is the model in which the variance follows a latent stochastic process. These models are called stochastic volatility model which was introduced in 1994 by Jacquier [10] and Shephard [19].

The empirical version of basic SV model in discrete time is:

$$y_t = \beta exp(h_t/2)\varepsilon_t,$$
$$h_{t+1} = \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t,$$
$$h_1 \sim N(\mu, \frac{\sigma^2}{1-\phi^2})$$

where y_t is the mean corrected return, h_t is the log volatility at time t which is assumed to follow a stationary process ($|\phi| < 1$). Compared to the ARCH/GARCH models, there are two shock terms, ε_t and η_t, for asset return and volatility respectively. They 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 autoregressive model of $\mathbf{h}$, such impact will influence $\mathbf{h}^t = (h_{t+1}, ..., h_n)$. These two stochastic innovations in SV model provides more flexibility compared to ARCH/GARCH models, but at the same time, it complicates estimation because $\mathbf{h}$ is now an unobserved stochastic process.

The volatility process $\mathbf{h}$ follows a mean reverting first order autoregressive model, with mean volatility level μ, volatility reverting persistency parameter ϕ and variance of volatility σ^2. We can denote the parameters for this volatility equation by $\theta = (\mu, \phi, \sigma)$.

1.5 Generalization of Stochastic Volatility model

1.5.1 SV model considering leverage effect

To correlate the two error terms in return and volatility can extend the basic SV model allowing it to incorporate a leverage effect[11]. According to the definition of leverage, that an increase in return will decrease volatility and vice versa, the correlation ρ between the two innovation terms should be negative.

Hence a SV model with leverage is given by:

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

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

1.5.2 Regime Switching Stochastic Volatility Models

1.5.2.1 Stochastic Volatility Model With Markov Switching

Lamoureux and Lastrapes (1990)[13]suggested that the documented high persistency of variance may have been overestimated because structural shifts in the market are not taken into account. Hamilton (1994)[6] included markov switching in ARCH /GARCH models, and in 1998, So [22] incorporated regime switching properties into stochastic volatility models. In their model:

$$h_{t+1} = \mu_{D_{t+1}} + \phi(h_t - \mu_{D_{t+1}}) + \sigma\eta_t \tag{1.10}$$

where

$$\mu_{D_{t+1}} = \gamma_1 + \sum_{j=2}^{K} \gamma_j I_{jt} \tag{1.11}$$

and I_{jt} is an indicator variable that equals 1 when $D_t \geq j$. D_t is unobserved discrete regime variable which follows a K state first order Markov Process. This model is denoted as MSSV(K), and bayesian methods are applied for inference.

1.5.2.2 Markov-switching Diffusion Model

Smith (2000) [20] compared the Markov switching Diffusion and the stochastic volatility models of short-term interest rate. Markov switching diffusion model differs from the MSSV(K) model in that the volatilities within each regime are constant. In our representation, the Markov Switching model is:

$$y_t = \sigma_{D_t} \varepsilon_t \tag{1.12}$$

where D_t is the particular regime at time t and σ_i is the standard deviation of y_t in that regime. Quasi maximum likelihood estimation is implemented and the author concludes that "Markov-switching or stochastic volatility, but not both, are needed to adequately to fit the data" and the "Markov-switching model is the best in terms of forecasting".

1.5.2.3 Stochastic Volatility Models with Markov Regime Switching State Equations

Pereira (2004) [9] generalized the MSSV(K) model and introduced a Stochastic Volatility Models with Markov Regime Switching State Equations(SVMRS) which can investigate properties of volatility, volatility persistency and smoothness.

This model allows both ϕ and σ to change and assumes two possible regimes:

$$h_{t+1} = \mu_{D_{t+1}} + \phi_{D_{t+1}}(h_t - \mu_{D_t}) + \sigma_{D_{t+1}} \eta_t \tag{1.13}$$

The same as in the Markov-switching Diffusion model, D_t is the unknown discrete regime indicator variable and Quasi maximum likelihood estimation is adopted in parameter inferences.

1.6 Inference Methods for the Hidden Markov Chain

The main challenge in making inference for the stochastic volatility models lies in making inference for the unobserved stochastic process **h**. Furthermore, incorporating regime switching into the SV models adds another layer of complexity as inference for **D** is now regaved, which makes model estimation even more complicated. Therefore efficient inference methods for hidden Markov chains are needed in our research. Detailed bayesian inference approaches for **h** will be described in Chapter 3, and in this section, I will introduce several popular approaches of inference for the hidden Markov chain in the regime switching stochastic volatility models.

1.6.1 Quasi Maximum Likelihood (QML) based method

Pereira [9] proposed a Quasi Maximum likelihood method using the Kalman filter in his SVMRS model.

In the SVMRS model, the observational equation $y_t = exp(h_t/2)\varepsilon_t$ is firstly transformed to a linear model $y_t^* = \log y_t^2 = h_t + \psi_t$, where ψ_t is $\log \chi^2$ distribution. Following Harvey, Ruiz and Shephard (1994)[8], they treated ψ_t as though it were from $N(0, \frac{\pi^2}{2})$, and maximized the resulting quasi-likelihood function.

According to the SVMRS model, given the knowledge about 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):

$$
\begin{aligned}
D_{t+1} = 0, D_t = 0, \quad &then \quad h_{t+1} = \mu_0 + \phi_0(h_t - \mu_0) + \sigma_0\eta_t \\
D_{t+1} = 0, D_t = 1, \quad &then \quad h_{t+1} = \mu_0 + \phi_0(h_t - \mu_1) + \sigma_0\eta_t \\
D_{t+1} = 1, D_t = 0, \quad &then \quad h_{t+1} = \mu_1 + \phi_1(h_t - \mu_0) + \sigma_1\eta_t \\
D_{t+1} = 1, D_t = 1, \quad &then \quad h_{t+1} = \mu_1 + \phi_1(h_t - \mu_1) + \sigma_1\eta_t
\end{aligned} \tag{1.14}
$$

The number of combinations will increase 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

$$v_t^{ij} = y_t^* - h_{t|t-1}^{ij} \tag{1.15}$$

and

$$f_t^{ij} = E[(h_t - h_{t|t-1}^{ij})^2 | D_t = i, D_{t-1} = j, I_{t-1}] + \sigma_\phi^2 \tag{1.16}$$

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

$$f(y_t^* | D_t = i, D_{t-1} = j, I_{t-1}) = \frac{1}{\sqrt{2\pi f_t^{ij}}} \exp \frac{-(v_t^{ij})^2}{2f_t^{ij}} \tag{1.17}$$

.

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

$$\boldsymbol{L}(\mathbf{y}^* | \boldsymbol{\theta}) = \sum_{t=1}^{T} \sum_{ij} \log[f(y_t^* | D_t = i, D_{t-1} = j, I_{t-1}) p(D_t = i, D_{t-1} = j | I_{t-1})] \tag{1.18}$$

which can be easily calculated. Numeric maximization of the loglikelihood function leads to the QML estimator of $\boldsymbol{\theta}$. Various starting values of $\boldsymbol{\theta}$ must be tried in order to avoid local maximization and typically a smoothing technique is applied to extract 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 the of Pereira [9].

1.6.2 Bayesian Method

Stochastic volatility models with Markov switching [22] are 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 their paper, So (1998)[22] 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

$$f(\mathbf{D}|\mathbf{y}, \mathbf{h}, \boldsymbol{\theta}) = f(D_n|\mathbf{y}, \mathbf{h}, \boldsymbol{\theta}) \prod_{t=1}^{n-1} f(D_t|\mathbf{y}, \mathbf{h}, \boldsymbol{\theta}, \mathbf{D}^{t+1}), \tag{1.19}$$

where $\mathbf{D}^t = (D_t, ..., D_n)$. Therefore 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 Markov property,

$$f(D_t|\mathbf{y}, \mathbf{h}, \boldsymbol{\theta}, \mathbf{D}^{t+1}) \propto f(D_{t+1}|D_t) f(D_t|\mathbf{y}, \mathbf{h}, \boldsymbol{\theta}) \tag{1.20}$$

and a discrete filter developed by Carter and Kohn (1994)[2] can be applied to evaluate $f(D_t|\mathbf{y}, \mathbf{h}, \boldsymbol{\theta})$. A simultaneous sampling of $\mathbf{D}$ is hence straightforward.

Chapter 2

Motivation

The regime switching stochastic volatility models introduced earlier can be improved in several aspects.

Firstly, all of the existing regime switching models are based on discrete time hidden Markov process $\mathbf{D}$. While these models are relatively simple, they tend to produce high frequency of transitions between regimes. Such frequent switches lead to over fitting and an inability to identify large shocks. For example, the current subprime crisis represents a large shock to the financial market. It has changed assets' volatility behaviors and should have a lasting impact. Retrospecting after 50 years, people will tend to relate the volatility performance changes to this crisis happened in 2008, rather than a lot of possible smaller shocks thereafter. For the purpose of connecting shifts with real events, a model which can relate major news event with volatility structural changes is required.

Secondly, the existing models are not flexible enough. In MSSV(K) model [22], only volatility mean level μ varies by regimes. In Markov Switching model [20], volatilites are assumed to be constant within each regime. For the SVMRS model [9] all volatility parameters μ, ϕ, σ^2 can be different, however only two possible regimes are considered in their framework. Empirical work calls for a much more flexible model because there are probably more than two regimes in the market, and volatility persistency should not be constrained to be identical under different circumstances.

The third area of improvement resides in extending inference techniques. Stochastic volatility models are notoriously difficult to estimate because the likelihood is not in closed form. Moreover, volatility process $\mathbf{h}$ is both unobserved and random. Adding a regime switching process $\mathbf{D}$ into SV model makes estimation even more complicated. Quasi-MLE [20][9] works fine in regime switching SV model inferences, but by augmenting parameter set (μ, ϕ, σ^2) with the unknown processes $\mathbf{h}$ and $\mathbf{D}$, a bayesian MCMC algorithm can be more efficient.

Based on these observations, we proposed a continuous version regime switching stochastic volatility model, where the total number of regimes K is not restricted to 2, (μ, ϕ, σ^2) can differ by regimes, and the hidden Markov chain $\mathbf{D}$ is a continuous time Markov chain with jumps and waiting times in between. We estimate parameters in the bayesian framework. Gibbs sampling, slice sampler, reversible jump sampler all played roles in our inference. Moreover, an offset mixture representation filter is applied to update the volatility process $\mathbf{h}$.

Our model succeeds in reducing the regime switching chances and at the same time kept the good attributes of regime switching volatility models. Moreover, by allowing all parameters to change by regime, our model can explain volatility persistency and smoothness property.

Chapter 3

Stochastic Volatility with Regime Switching - Statistical Inference

The canonical regime switching stochastic volatility model is:

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

where y_t is the response variable, and h_t is the unobserved log volatility at time t which is assumed to follow a stationary ($|\phi| < 1$) first order autoregressive process. D_t is the hidden regime at time t, ε_t and η_t are uncorrelated standard normal variables. Denote $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma})$.

The decomposition of the joint posterior density according to Baye's rule is:

$$f(\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}, \mathbf{h}, \mathbf{D}|\mathbf{Y}) \propto f(\mathbf{Y}|\mathbf{h})f(\mathbf{h}|\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}, \mathbf{D})f(\mathbf{D}|\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma})f(\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}) \tag{3.2}$$

The following sections are organized in two steps. First, we review inferences for the simple stochastic volatility model and then we will review inferences for the hidden Markov chain in the second part.

3.1 Inference for basic SV models

In the basic SV model, the augmented parameter set is $(\mu, \phi, \sigma, \mathbf{h} = (h_1, h_2, ..., h_n))$ and $\theta = (\mu, \phi, \sigma)$ is not in boldface because each parameter is a scaler in this model.

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

The likelihood function, $f(\mathbf{Y}|\theta) = \int f(\mathbf{Y}|\mathbf{h})f(\mathbf{h}|\theta)d\mathbf{h}$, has no analytical form which complicates the inference procedure. Without loss of generality, we can assume that $\beta = 1$, that is, $\mathbf{Y}$ has been mean corrected.

The decomposition of the joint posterior density for this model can be viewed as a hierarchical model structure as in the paper by Jacquier (1994) [10]:

$$\pi(\mathbf{h}, \theta|\mathbf{y}) = f(\mathbf{y}, \mathbf{h}, \theta)/f(\mathbf{y}) \sim f(\mathbf{y}|\mathbf{h})f(\mathbf{h}|\theta)f(\theta) \tag{3.3}$$

The two blocks of parameters, θ and $\mathbf{h}$, can be sampled separately using the Gibbs sampler.

3.1.1 Sampler from full conditional distribution of θ

Inference for parameters θ can be made by integrating out $\mathbf{h}$ in eq. 3.3 and then sampling from the marginal distribution. We follow the method proposed by Kim (1998) [12] for obtaining these marginal densities.

3.1.1.1 Inference for σ

Conjugate prior $IG(\sigma_r/2, S_\sigma/2)$ is assumed for σ^2 and the posterior distribution for σ^2 is:

$$f(\sigma^2|\mathbf{y}, \mathbf{h}, \phi, \mu) \sim IG(\frac{N + \sigma_r}{2}, \frac{S_\sigma + \sum_{i=1}^{N} r_i^2}{2}) \tag{3.4}$$

where IG stands for inverse gamma distribution, with shape and scale parameters, σ_r and S_σ. r_i is the residual from volatility equation, that is, $r_i = h_{i+1} - \mu - \phi(h_i - \mu)$.

3.1.1.2 Inference for ϕ

Let $\phi = 2\phi^* - 1$, where ϕ^* is distributed with a Beta distribution with parameters a, b. The prior distribution for ϕ is

$$f(\phi) \propto \{\frac{1+\phi}{2}\}^{a-1}\{\frac{1-\phi}{2}\}^{b-1} \tag{3.5}$$

The full conditional distribution of ϕ is proportional to $f(\phi)f(\mathbf{h}|\mu, \phi, \sigma)$ where $f(\mathbf{h}|\mu, \phi, \sigma)$ is a normal distribution. Metropolis Hasting algorithm is used to sample from this full conditional distribution. For a discussion of the Metropolis -Hastings algorithm, see [12].

3.1.1.3 Inference for μ

A diffuse prior on μ is assumed, hence, the full conditional distribution for μ is

$$f(\mu|\mathbf{h}, \phi, \sigma) \sim N(\hat{\mu}, \sigma_\mu^2) \tag{3.6}$$

where

$$\hat{\mu} = \sigma_\mu^2\{\frac{1-\phi^2}{\sigma^2}h_1 + \frac{1-\phi}{\sigma^2}\sum_{i=1}^{N-1}(h_i - \phi h_{i-1})\} \tag{3.7}$$

$$\sigma_\mu^2 = \sigma^2\{(n-1)(1-\phi^2) + (1-\phi^2)\}^{-1} \tag{3.8}$$

3.1.2 Sampler from full conditional distribution of h

3.1.2.1 Sample h one at a time

Sampling from the distribution for $(\mathbf{h}|\mathbf{y},\theta)$ is more complicated than sampling from the distribution for θ. For example, Jacquier (1994)[11] considered the conditional distribution of each element of **h**, or

$$p(h_t|\mathbf{h}_{-t},\mathbf{y},\theta) = p(h_t|h_{t-1},h_{t+1},\theta,\mathbf{y}) \tag{3.9}$$

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

The explicit full conditional distribution for each element in **h** is

$$p(h_t|h_{t-1},h_{t+1},\theta,\mathbf{y}) \propto h_t^{-0.5} exp(-0.5y_t^2/h_t)1/h_t exp-(lnh_t - \mu_t)^2/(2\sigma^2), \tag{3.10}$$

where $\mu_t = (\mu(1-\phi) + \phi(lnh_{t+1} + lnh_{t-1}))/(1+\phi^2)$ and $\sigma^2 = \sigma_\eta^2/(1+\phi^2)$.

There are a number of competing strategies for sampling **h**, see for example, Shephard (1993), Jacquier, Polson and Rossi (1994)[11], Shephard and Kim (1998)[12], Geweke (1994)[5] and Shephard and Pitt (1997) [17]. Jacquier, Polson and Rossi (1994)[11] proposed using an accept/reject Metropolis algorithm while in most other cases, acceptance-rejection sampling methods were enough.

3.1.2.2 Sample h simultaneously

Since $\{h_t\}$ are highly correlated, the initial updating schemes in the literature were not efficient in order to overcome this correlation. Kim (1998)[12] suggested using an offset mixture representation to approximate the distribution of the error for h_t, which allows for filtering based methods which can generate **h** simultaneously from a multivariate distribution. To be explicit,

$$y_t^* = h_t + z_t \tag{3.11}$$

where $y_t^* = log(y_t^2) + c$ and

$$f(z_t) = \Sigma_{i=1}^L q_i f_N(z_i|m_i - 1.2704, \nu_i^2). \tag{3.12}$$

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(1998)[12], $L = 7$ 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 indicator variable s_t such that

ω	$Pr(\omega = i)$	m_i	σ_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 $log\chi^2$

$$z_t|s_t = i \sim N(z_i|m_i - 1.2704, \nu_i^2), Pr(s_t = i) = q_i. \tag{3.13}$$

Augmenting the previous parameter set with $\mathbf{s}$, there are now three blocks of parameters for this model: $(\theta, \mathbf{h}, \mathbf{s})$ and now $\mathbf{s}$ can be sampled independently from its full conditional distribution given below,

$$Pr(s_t = i|y_t^*, h_t) \propto q_i f_N(y_t^*|h_t + m_i - 1.2704, \nu_i^2). \tag{3.14}$$

Because $\mathbf{y}^*|\mathbf{s}, \theta, \mathbf{h}$ is a Gaussian time series, it can be placed into a state space form associated with the Kalman filter. In this case, it is possible to sample from the multivariate Gaussian distribution $\mathbf{h}|\mathbf{y}^*, \mathbf{s}, \theta$ simultaneously. A description of the sampling algorithm used by Jong and Shephard (1995)[3] is given in Appendix A. Kim (1998) [12] also proposed an integration sampler which can improve the mixing behavior between $\mathbf{h}$ and θ.

3.2 Inference with regime switching

We now turn to models that accommodate different regimes within the stochastic volatility framework.

3.2.1 Inference for $\boldsymbol{\theta}$

For each regime i of K possible regimes, there is one corresponding set of parameters that governs the resulting stochastic volatility for that regime, μ_i, ϕ_i and σ_i. Conditional on the hidden regime chain $\mathbf{D}$, the inference for these parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma})$ is straightforward. The information can be collected from each regime so that basic SV inference can be performed respectively.

To avoid the problem with identification, we specified an order based on the states of the underlying chain, e.g., $\mu_1 < \mu_2 < ... < \mu_K$ for $\boldsymbol{\mu}$. The intuition behind this is that we believe that regimes are differentiated by average volatility levels, which is a common specification in the literature of regime switching. Although this is a standard assumption, it is easy enough to empirically test other alternatives; for example, is it the persistency parameter $\boldsymbol{\phi}$ or the variance of volatility $\boldsymbol{\sigma}$, rather than $\boldsymbol{\mu}$ the parameters which drive regime shifts? We addressed this question by setting the order for $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}$ separately, or "anchoring $\boldsymbol{\phi}$" or "anchoring $\boldsymbol{\sigma}$".

When the parameter orders are specified, inferences for the ordered parameter is not as straightforward which allowed us to use the slice sampler (Neal 2003)[15].

3.2.1.1 Inference when $\boldsymbol{\mu}$ is anchored

The full conditional distribution of μ_i is $N(\hat{\mu_i}, \sigma_{\mu,i})$ with upper bound and/or lower bound. The slice sampling algorithm that we used is as follows. First note that the density of a normal distribution can be written as

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(x-\mu)^2}{2\sigma^2} \tag{3.15}$$

Step 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.

Step 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).

Step 3: Sample $\mu_i \sim Unif(\max(LB_i, \mu_{i-1}), \min(UB_i, \mu_{i+1}))$.

3.2.1.2 Inference when $\boldsymbol{\sigma^2}$ or ϕ are anchored

The full conditional distribution of $\boldsymbol{\sigma}^2$ in eq. 3.4 is an inverse gamma distribution. The density of inverse gamma distribution is

$$f(x) \propto x^{-\alpha-1} \exp(\frac{-\beta}{x}) \tag{3.16}$$

There are two parts that involve x in eq. 3.16, and they can be treated separately in deriving the slice sampler routine.

Step 1: assume the current value for σ_i^2 is x. Sample $u_1 \sim Unif(0, x^{-\alpha-1})$; then sample $u_2 \sim Unif(0, \exp(\frac{-\beta}{x}))$.

Step 2: solve the equation $y_1^{-\alpha-1} = u_1$ to get a lower bound y_1, LB_i; then solve $\exp(\frac{-\beta}{y_2}) = u_2$ to get an upper bound y_2, UB_i.

Step 3: Sample $\sigma_i^2 \sim Unif(\max(LB_i, \sigma_{i-1}^2), \min(UB_i, \sigma_{i+1}^2))$;

The same idea is applicable to $\boldsymbol{\phi}$ although its full conditional distribution is not in a closed form.

3.2.2 Inference for D

In the literature, the hidden regimes $\{D_t\}$ in stochastic volatility models are all regarded as a discrete Markov Chain. Such stipulation gives rise to frequent structural shifts. Our purpose is to identify longer periods of stochastic volatility as we want to relate these structural shifts to certain news events. Therefore we modeled $\{D_t\}$ as a continuous-time Markov chain, see Liechty and Roberts (2001)[14].

$\{D_t\}$ is modeled in terms of jump chain $(i_0, i_1, ...)$ and waiting times between jumps $(t_0, t_1, ...)$, which we call intervals. For simplicity, the waiting times are modeled as exponential distributions with average waiting time parameter λ_i. This parameter can be different for different regimes. More general interval distributions, Gamma for example, can be explored in the future. The parameters for the stochastic volatility model are augmented with $\mathbf{P}$, which is the jump transition matrix.

3.2.2.1 Inference for $\mathbf{P}, \lambda$

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

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

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 distribution a priori.

The transition matrix $\mathbf{P}$, together with the inverse of average waiting time $\boldsymbol{\lambda}$, give rise to an unconditional transition matrix $\mathbf{Q}$ for $\mathbf{D}$ (Liechty and Roberts 2001)[14].

$$\begin{aligned} Q_j &= p_{lj} \times \lambda_l \quad if \quad l \neq j, \\ &= 1 - \lambda_l \quad if \quad l = j \end{aligned} .$$

The waiting time parameter, $\boldsymbol{\lambda}$, can be updated using the Gibbs sampler as the prior for $\boldsymbol{\lambda}$ is a Gamma distribution which results in a conjugate full conditional density.

3.2.2.2 Algorithm to update D

The hidden Markov Chain $\mathbf{D}$ can be characterized by the number of intervals and the starting point of each interval. Because the number of parameters needed to describe a particular realization of $\mathbf{D}$ will change as the number of intervals changes, a Reversible Jump Metropolis-Hastings updating schemes is used to sample $\mathbf{D}$. The details are clearly stated in Liechty and Roberts (2001)[14].

There are two parts of the sampling schemes that are worth noting. First, because the last interval is censored, when calculate the likelihood for a realization of $\mathbf{D}$, the probability for it is the cumulative density $p(X \geq x)$, where x is the length of the last interval.

Consideration has to do with finding a reasonable starting value for $\mathbf{D}$. Following is the strategy that we followed in practice.

Step 1: start with a basic stochastic volatility model. Generate a large number of iterations of the basic SV model which can provide a reasonably good estimate of $\mathbf{h}$ and $\mathbf{s}$.

Step 2: $\mathbf{D}$ is anchored by $\boldsymbol{\mu}$. Set the initial values of $\boldsymbol{\mu}$ by dividing up the posterior distribution of $\boldsymbol{\mu}$ from the basic SV model.

Step 3: Keep $\mathbf{h}$ and $\boldsymbol{\mu}$ fixed to these initial values and update $\mathbf{D}$ and other parameters for a large number of iterations.

Step 4: finally run the full algorithm with these initial values.

Chapter 4

Empirical Results and Conclusions

We applied our continuous time regime switching SV model to several financial datasets: $S\&P500$, Federal Rate, Exchange Rate, and Stock Price. In the following we compare the regime switching SV model with the benchmark SV model using a range of model comparison tools. The highlights of these results include:

- The regime switching SV model outperforms the benchmark model in both model fitting and forecasting.
- The regime switching SV model identifies larger intervals and less frequent switches when compared to discrete-time regime switching models, which potentially relate these switches to historical news events.
- The regime switching SV model is very flexible in that we can specify any value for K ($K >= 2$) as the number of underlying structures (regimes). It also allows the parameters $(\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma})$ to differ by regimes.
- Using model selection tools, we can find the best model out of all possible candidate models including different number of regimes and different anchoring schemes. This best model not only has statistical significance, but also can explain possible econometric/financial phenomena.
- Based on our empirical work, we conclude that market structures should be determined by anchoring the average volatility level $\boldsymbol{\mu}$, rather than $\boldsymbol{\phi}$ or $\boldsymbol{\sigma}$.

4.1 S&P 500 daily data

We will use S&P 500 daily data to illustrate our method. It is ten years' $S\&P$ 500 daily close value data from 02/27/1992 to 02/27/2002. The plot of raw data is shown in Figure 4.1. This figure suggests an evident inconsistency in data structures. For the first half, the values are at a lower level, and very stable, while in the second half, the closing values go up, and at the same time, they become much more volatile.

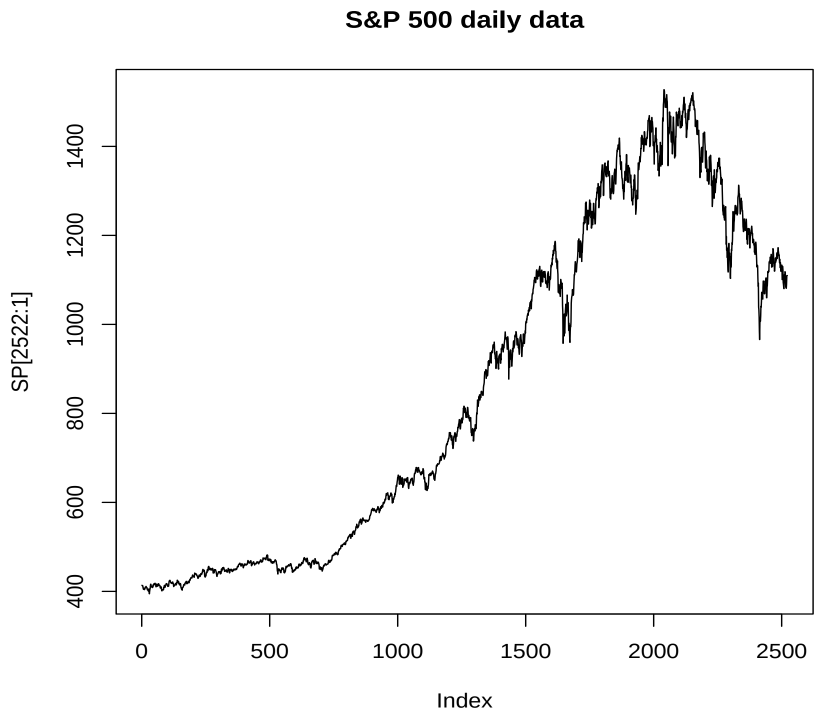


Figure 4.1. SP 500 original data

We transformed the data by taking relative difference of the original $S\&P500$ data, $y_i = \frac{x_{i+1}-x_i}{x_i}$. To avoid possible rounding error problems, the transformed series $\mathbf{y}$ is multiplied by 100, which preserves the property of the series. Zero values in the data are assigned a very small value, 0.01.

The plot of transformed data is shown in Figure 4.2. From these point on, I will refer to the transformed series $\mathbf{y}$ as the $S\&P500$ data and all modeling and forecasting are based on the transformed data. Both the benchmark basic SV model and our Regime Switching SV models are applied to this dataset .

4.1.1 Regime Switching SV model

We assumed two possible regimes ($K = 2$) for $S\&P500$ data by visual diagnosis. Within this two regime framework, there are twelve possible regime switching stochastic volatility models. They are :

Anchor $\boldsymbol{\mu}$:

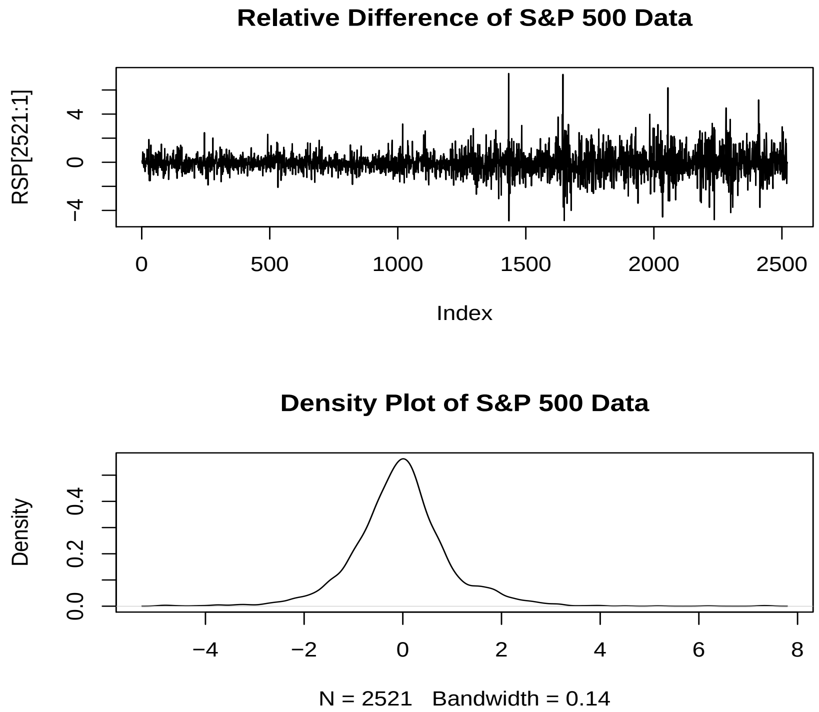


Figure 4.2. SP 500 differenced data

Market structural differences are determined by the average volatility levels, or

$$\mu_1 < \mu_2 \tag{4.1}$$

1. $\phi_1 = \phi_2$ and $\sigma_1^2 = \sigma_2^2$.
2. $\phi \neq \phi_2$, but $\sigma_1^2 = \sigma_2^2$.
3. $\sigma_1^2 \neq \sigma_2^2$ but $\phi_1 = \phi_2$.
4. $\phi_1 \neq \phi_2$ and $\sigma_1^2 \neq \sigma_2^2$.

Anchor ϕ:

Market structural differences are determined by how fast volatility reverts to its average level.

$$\phi_1 < \phi_2 \tag{4.2}$$

1. $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$.
2. $\mu \neq \mu_2$, but $\sigma_1^2 = \sigma_2^2$.
3. $\sigma_1^2 \neq \sigma_2^2$ but $\mu_1 = \mu_2$.
4. $\mu_1 \neq \mu_2$ and $\sigma_1^2 \neq \sigma_2^2$.

Anchor σ^2:

Market structural differences are determined by the variance of volatilities.

$$\sigma_1^2 < \sigma_2^2 \tag{4.3}$$

1. $\mu_1 = \mu_2$ and $\phi_1 = \phi_2$.
2. $\mu \neq \mu_2$, but $\phi_1 = \phi_2$.
3. $\phi_1 \neq \phi_2$ but $\mu_1 = \mu_2$.
4. $\mu_1 \neq \mu_2$ and $\phi_1 \neq \phi_2$.

4.1.2 Model Selection

For the regime switching SV models, the determination of the total regime number (K) is difficult and sometimes arbitrary. For the $S\&P500$ data, as mentioned before, we assigned $K = 2$ based on empirical evidence. However K can take other values as well. This challenge requires formal model choice tools.

4.1.2.1 Maximum Likelihood and Reversible Jump Model Selection

The simple Maximum Likelihood (ML) comparisons will recommend a model that gives the largest likelihood. A reversible jump model selection algorithm improves upon ML by penalizing models with more parameters in an appropriate manner. Using the reversible jump algorithm [21], the posterior probability for each model will be calculated, and the model with the highest posterior probability is selected.

4.1.2.2 Out of Sample Test

Forecasting is an alternative model selection criteria for these analysis. We adopted Out-of-Sample prediction summaries which contrast forecasting accuracy of each model. The Out-of-Sample test is implemented in the following way:

- We divide the historic data into two subsets: a "fit" period followed by a "test" period. The length of the "test" period varies from the last 100 to the last 200 elements of the $S\&P500$ dataset. Denote the final element of the "fit" period by T_{ff} and the first element of the "test" period by T_{op}.

- Add a forecasting step in each MCMC estimation cycle of the SV models.

  The forecasting steps for the "test" period are as follows:

  Step 1: determine the hidden regime D_t for each time point;

  Step 2: calculate the predicted return values $\hat{y}_t$;

  Step 3: take absolute differences $|\hat{y}_i - y_i|$ and squared differences $(\hat{y}_i - y_i)^2$ and store them.

- After MCMC burn-in and sampling iterations, take average over the absolute differences and squared differences. The summary statistics are called the Out-of-sample-Mean-Absolute-Difference (OMAD) and the Out of sample Out-of-sample-Root-Mean-Squared Error (ORMSE). A model with small OMAD and ORMSE values is desirable.

4.1.3 Empirical Results for $S\&P500$ Data

Parameter estimation and model selection results are shown in Table 4.1. The best model can be identified using the model choice tools. Several interesting conclusions can be drawn from these results:

- The estimates of $\boldsymbol{\mu}$ are consistent, regardless of the anchoring scheme as long as $\boldsymbol{\mu}$ is allowed to vary by regimes. This fact holds regardless of whether the other parameters differ by regimes.
- Models that do not allow $\boldsymbol{\mu}$ to vary are not able to capture regime shifts and structural changes. For example, in models 5 and 7 , and in models 9 and 11, where $\mu_1 = \mu_2$, the parameter estimates are all very close between the different regimes. Moreover, models that anchor μ perform better than the competing models that do not.
- From the above observations we can conclude that volatility's reverting means $\boldsymbol{\mu}$ determine regime(structural) differences. Although $\boldsymbol{\phi}$ and $\boldsymbol{\sigma^2}$ can differ by regimes, they are not the driving force of the structural differences. In the future, only models that allow $\boldsymbol{\mu}$ to differ by regime are referred to as the regime switching stochastic volatility model.
- As noted in the literature, spurious high persistency in models where $\mu_1 = \mu_2$ is evident. When regime switching is allowed in the model, the persistency parameter, ϕ, decreased substantially from over 0.95 to around 0.5.
- In the low volatility regime (regime 1), ϕ is smaller and σ^2 is larger when compared with regime 2. The financial interpretation of this result is that when asset volatility is at a low level, larger sudden moves from the average volatility level are possible but these moves will tend to last for shorter periods.
- Compared to models without regime switching, the regime switching models do a better job at explaining the phenomena and predicting future price movements.

Using the model choice criterions, all of the models where $\boldsymbol{\mu}$ is anchored are better than models where $\boldsymbol{\mu}$ is not anchored. Among these models, model 3, where $\boldsymbol{\mu}$ is anchored and $\boldsymbol{\sigma^2}$ change, has the highest posterior probability, while the out-of-sample prediction criteria prefers model 4, where $\boldsymbol{\mu}$ is anchored and both $\boldsymbol{\phi}$ and $\boldsymbol{\sigma^2}$ change.

No	Model	μ	ϕ	σ^2	maxlike	postprob	OMAD	ORMSE
1	Anchor μ only	-1.33 0.22	0.27	0.57	-2774.12	0	0.75	5.35
2	Anchor μ, ϕ change	-1.42 0.16	0.26 0.58	0.38	-2747.22	0.0123	0.70	5.05
3	Anchor μ, σ change	-1.47 0.20	0.19	0.67 0.39	-2735.57	0.9872	0.65	3.09
4	Anchor μ, σ ϕ change	-1.38 0.30	0.02 0.87	0.51 0.09	-2830.1	0	0.61	3.05
5	Anchor ϕ only	-0.42	0.96 0.97	0.07	-2885.23	0	1.03	11.4
6	Anchor ϕ, μ change	-1.39 0.20	0.27 0.54	0.40	-2741.42	0.0004	0.66	3.4
7	Anchor ϕ, σ change	-0.42	0.92 0.97	0.21 0.04	-2933.44	0	0.97	6.45
8	Anchor ϕ, σ μ change	-1.16 -0.15	0.84 0.96	0.33 0.04	-2897.21	0.0	0.97	5.89
9	Anchor σ only	-0.44	0.94	0.11 0.14	-2902.39	0	1.11	11.6
10	Anchor σ , μ change	0.27 -1.2	0.85	0.09 0.11	-2917.45	0	0.66	3.23
11	Anchor σ , ϕ change	-0.42	0.96 0.95	0.07 0.09	-2926.37	0	1.06	12.03
12	Anchor σ , ϕ μ change	-1.22 0.26	0.86 0.89	0.08 0.09	-2934.07	0	0.56	2.03
13	Basic SV	-0.38	0.987	0.14	-2979.6	0	0.75	4.12

Table 4.1. SP 500 Parameter Estimates and Model Comparions

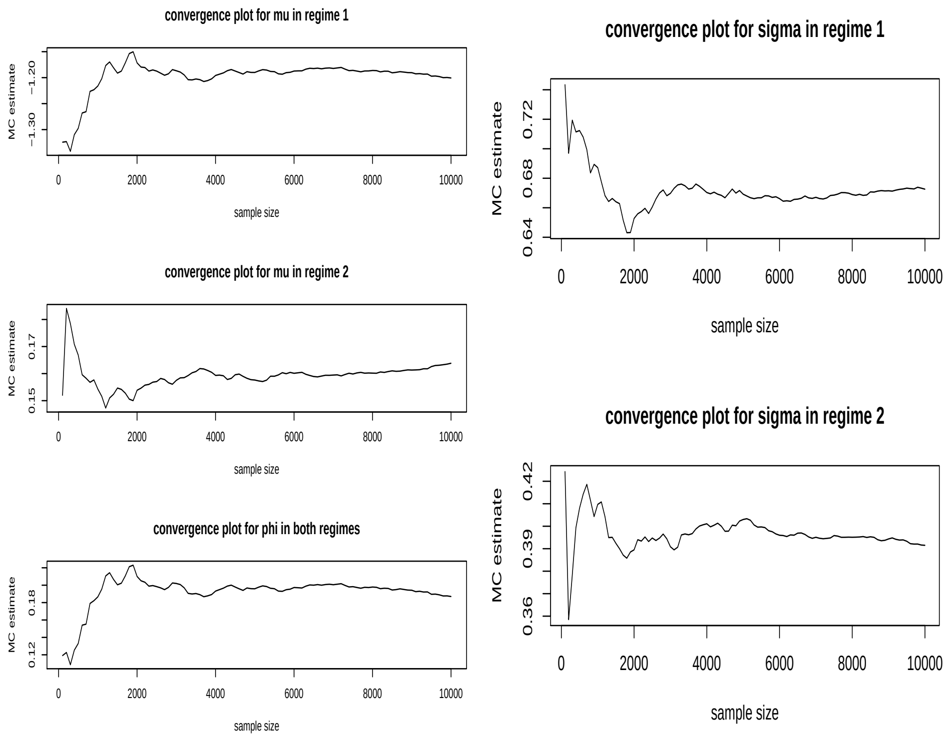
4.1.4 Visual Diagnoses

In addition to model choice criteria, we can provide several ad hoc visual diagnosis for assessing model fit. Several strategies are described below; we illustrate these summaries with results from model 3 (except when explicitly specified).

- **Convergence Plot**: This plot shows how Monte Carlo estimates change with an increase in the number of iterations. We ran the MCMC sampler for 10,000 burn-in iterations followed by 10,000 sampling iterations for each model. The parameter estimators appear to have converged as shown in Figure 4.3.

- **Posterior Regime Plot**: Figure 4.4 reports the posterior probability of being in each regime. The first panel is the plot of the data $\mathbf{y}$. The probability of being in a low volatility regime (regime 1) and a high volatility regime (regime 2) are displayed in the second and third panel respectively. Clearly the regime switching SV model captures the different volatility patterns of the $S\&P500$ dataset.

- **Percentiles for In-Sample Forecast Plot**: In-sample fit can be assessed by observing that $y_i \sim N(0, \exp(h_i))$, and for each sweep of the MCMC sampler, calculating the per-



(a)Convergence for μ and ϕ (b) Convergence for σ^2

Figure 4.3. SP 500 Convergence Plot for Model 3

centiles associated with $\mathbf{h}$. A posterior mean of the percentages are then plotted in Figure 4.5. The red lines are the 95th and 5th percentiles and the green lines are the 25th and 75th percentiles.

The first panel in Figure 4.5 presents percentiles for model 1, where $\mu_1 < \mu_2,\ \phi_1 = \phi_2,\ \sigma_1^2 = \sigma_2^2$. The second panel shows percentiles for model 4, where $\mu_1 < \mu_2,\ \phi_1 \neq \phi_2,\ \sigma_1^2 \neq \sigma_2^2$ while the third one corresponds to the basic SV model.

As described in chapter 1, ϕ reflects the volatility reverting speed. Consistent with our numerical results in Table 4.1, the basic SV model has a spuriously high volatility persistency [13] which is evident in the third panel of Figure 4.5. Moreover, a comparison of the percentile smoothness in panel 1 and panel 2 verified our previous conclusion that if different regimes do not have the same persistency parameter, the volatility is more persistent in the high volatility regimes. A similar type of in-sample visual assessment is the Absolute In-sample Plot. In this plot, we take the posterior means of the volatility $\hat{\mathbf{h}}$, and compares $\exp(\hat{\mathbf{h}}/2)$ (in green) with the absolute values of $\mathbf{y}$ (in black), see Figure 4.6. This plot gives a similar visual summary as the summary given in the percentile plot.

- **Model Realization Plot**: In this plot, we plug the posterior estimators $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\sigma}^2})$ into

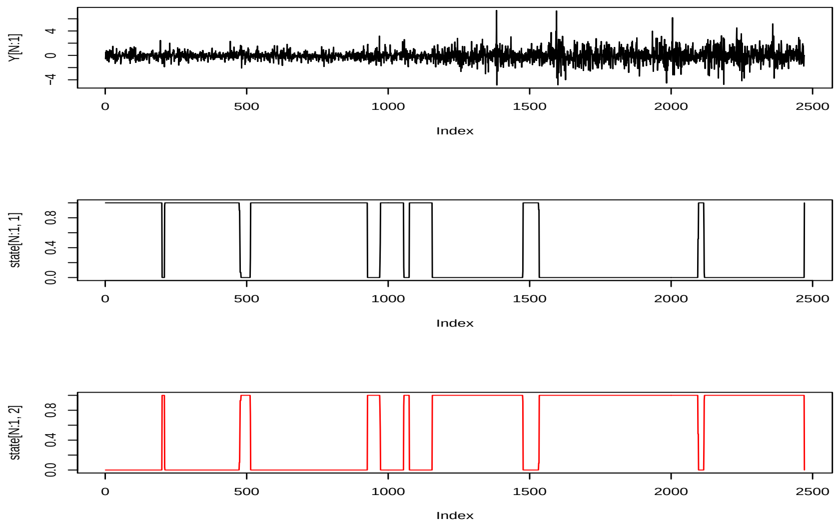


Figure 4.4. SP 500 Posterior Regime Plot

the stochastic volatility models and generate one realization of the data $\hat{\mathbf{y}}$. We then plot $\hat{\mathbf{y}}$ (in green) versus their true values $\mathbf{y}$ (in black), as shown in Figure 4.4. To generate the realization of $\hat{\mathbf{y}}$ for the regime switching SV model, we utilized the posterior regime probability as follows:

$$\begin{aligned} \hat{y}_t &\sim N(0, \exp(\hat{h}_t/2)) \\ \hat{h}_t &= p_{1t}(\hat{\mu}_1 + \hat{\phi}_1(h_{t-1} - \hat{\mu}_1) + \hat{\sigma}_1\eta_t) + p_{2t}(\hat{\mu}_2 + \hat{\phi}_2(h_{t-1} - \hat{\mu}_2) + \hat{\sigma}_2\eta_t) \end{aligned} \quad (4.4)$$

where p_{1t} is the posterior probability of being in regime 1 at time t.

The advantage of our model over the basic SV model is clear in these figures. For the basic SV model, the more volatile period is underestimated while the less volatile period is over estimated. This is because the model confined the parameters to be the same over two different structures, which is a stretch.

4.2 Other Applications

We have applied the same approach and diagnosis to many other financial datasets: US Dollar/British Pound exchange rate, federal interest rate, and IBM stock price. In each of these cases, there is clear evidence of regime switching structures. Figure 4.8 is an example of a model realization plot for the US Dollar/British Pound exchange rate, where we assumed $K = 3$. The second panel in this figure presents the density plot of true $\mathbf{y}$ (in black) and their estimations (in red). Both plots suggest the proposed model adequately captures the dynamics of the data.

Although we did not include all the results in this document, it is not surprising that different datasets, for example exchange rate, federal rate and stock prices, exhibited different patterns of behavior. Clearly from this set of empirical studies, we do need a range of regime switching models to address different types of financial data. Therefore beyond the general claim that financial data exhibit regime switching, we do not draw a general conclusion as to which regime switching model is the best. The proposed model is flexible, and comparison methods are straightforward, so it is not difficult to choose the best fitting model for a particular set of data.

4.3 Conclusion for Univariate Regime Switching Stochastic Volatility Model

The literature for modeling volatility is immense. A good model should help control risk, minimize loss, and serve as a basis for a better derivative pricing. Most empirical studies are based on the ARCH/GARCH framework as it is relatively easier to calibrate. However, the stochastic volatility model is more flexible and meaningful because it involves two sperate random terms for return and volatility respectively. It is an emerging and increasingly more popular area for research because of the additional insights it brings.

The basic stochastic volatility model is well established. A variety of inference methods, including MLE, EM algorithm, Quasi-ML and bayesian analysis, are present in the literature. In the bayesian framework, the most popular inference algorithm for the unobserved log volatilities $\mathbf{h}$ appears to be the offset mixture representation and Kalman filter proposed by Kim (1994)[12]. This approach is applicable to extensions of stochastic volatility models as long as the extended model can be transformed to a state space form. Stochastic volatility model with leverage effects should be excluded because it will lose important shock correlation information when transformed.

In the basic SV models, volatilities evolve in a diffusive way, which does not allow for abrupt changes in the market structures. However, in 1998 for example, the Asian Financial Crisis may have a tremendous impact on the financial market, which changed the volatility structures. For a graphical illustration of this change in the $S\&P$ 500 data, you can refer to the Figure 4.9. In this case, the basic SV model will be a poor fit, and it is reasonable to allow the model parameters coming from different regimes, with transitions between regimes governed by a hidden Markov chain. Although this is not a new idea, for example, Hamilton[6] proposed the switching regime ARCH/GARCH models, So [22], Smith [20], and Pereira [9] all introduced switching regimes in the stochastic volatility framework, all the existing regime switching models are based on discrete hidden regimes chains, which tend to give rise to too many transitions between regimes.

We proposed a continuous time regime switching model, which tend to result in fewer switches, which in turn are more likely to be related to certain news events. In our model, the continuous time hidden regime process consists of jumps between regimes and waiting times which have exponential distributions. Reversible jump MCMC method are used to update the hidden Markov chain.

The proposed model is very flexible as all the parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma})$ in the volatility equation are allowed to differ by regimes. Based on various model comparison and diagnosis methods, we conclude that the main driver of different regimes arise from differences in the mean volatility levels, $\boldsymbol{\mu}$. Moreover, for a more accurate characterization of each regime, an identification of whether different regimes share ϕ and σ^2 or not is needed. On the other hands, although people tend to describe market as only good or bad, there possibly exist more states in between. Therefore the total number of regimes, K, should not be restricted to two. In our model K is a random variable which is determined using model choice tools.

The proposed regime switching stochastic volatility model was applied to several financial datasets. The data we used include daily $S\&P500$ index, foreign exchange rate, federal funds rates and IBM stock prices. Different model comparison methods all confirmed that the proposed model outperforms the basic stochastic volatility model both in terms of model fitting and forecasting accuracy. These model choice tools are augmented by a set of ad hoc visualization tools that we have developed.

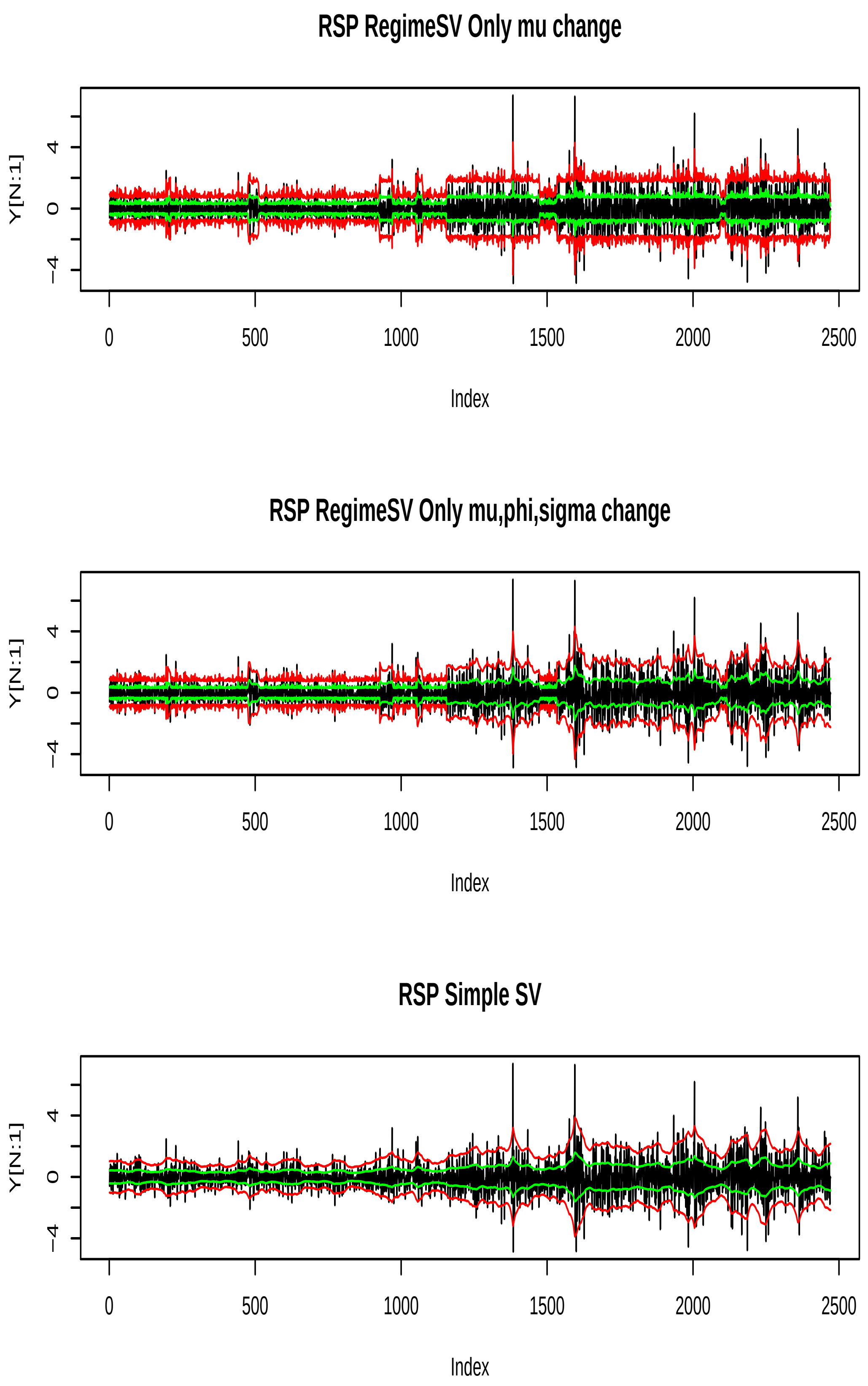


Figure 4.5. SP 500 Percentile-in-sample Plot

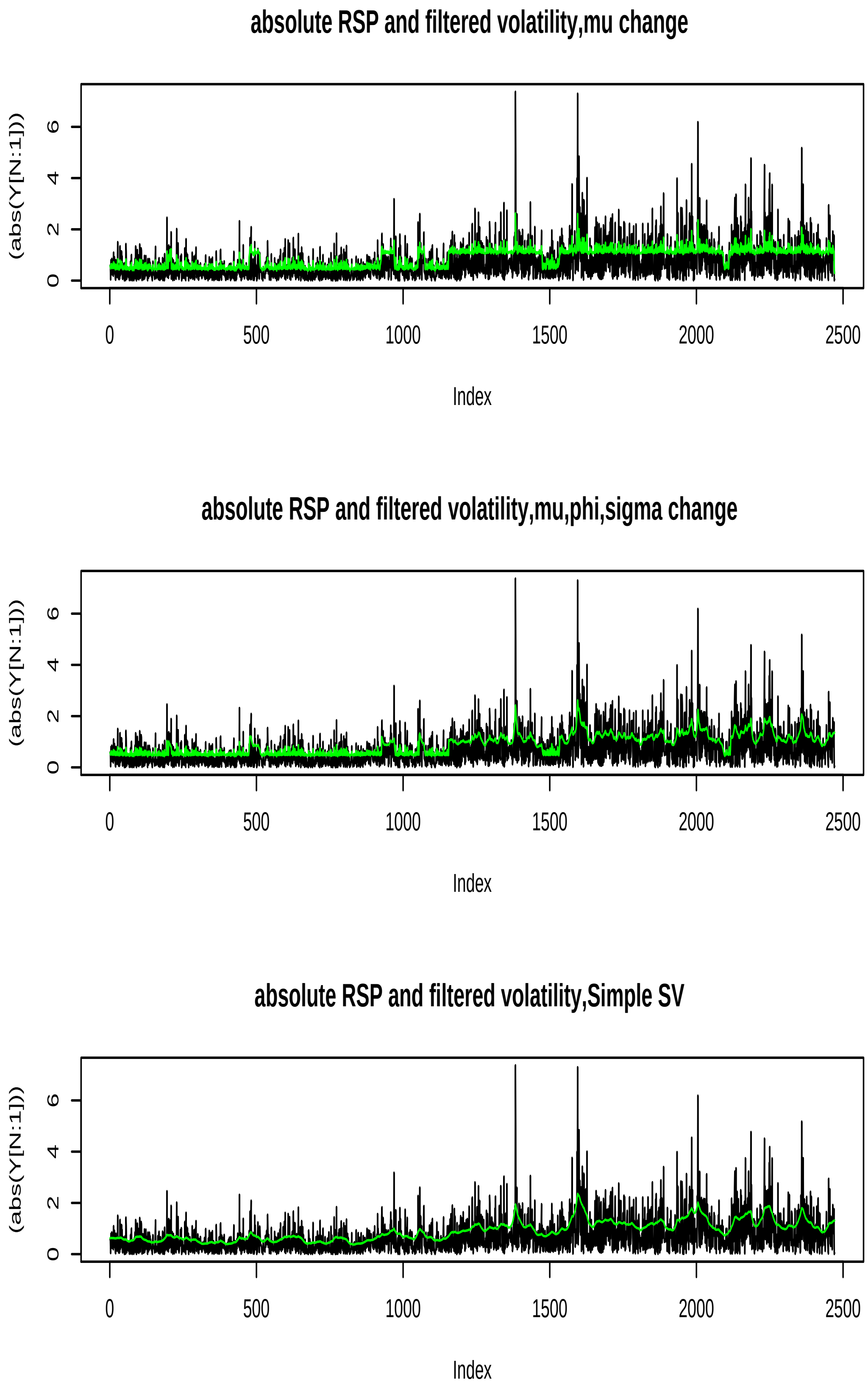


Figure 4.6. SP 500 Absolute In-sample Plot

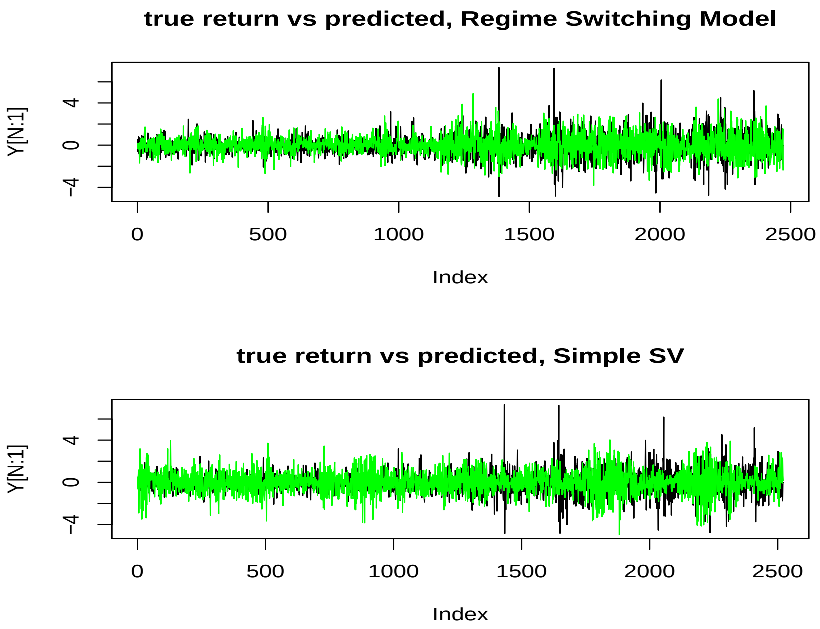


Figure 4.7. SP 500 Model Realization Plot for the Regime Switching SV model and the Basic SV model

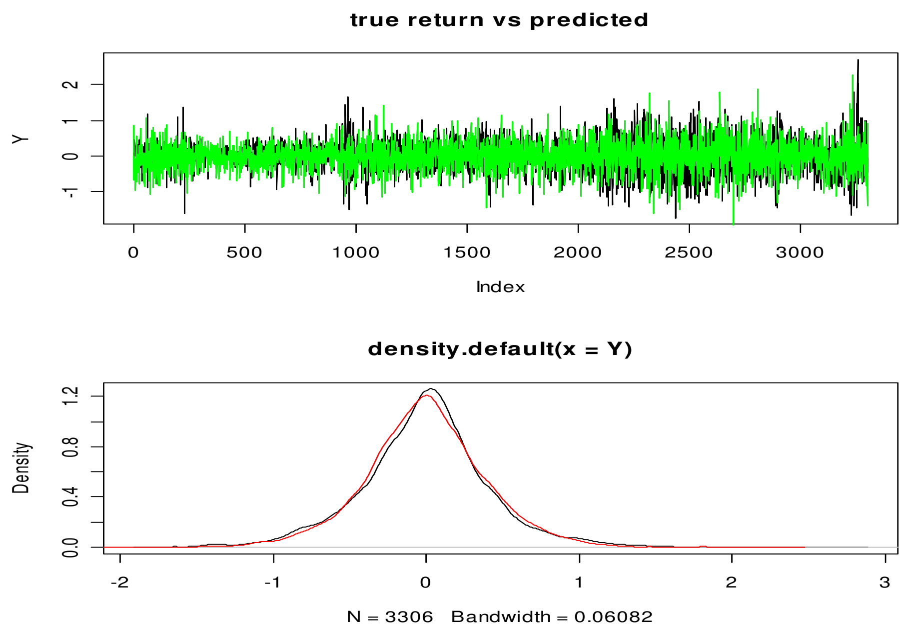


Figure 4.8. Exchange Rate Realization Plot and Density Plot by the Regime Switching SV model

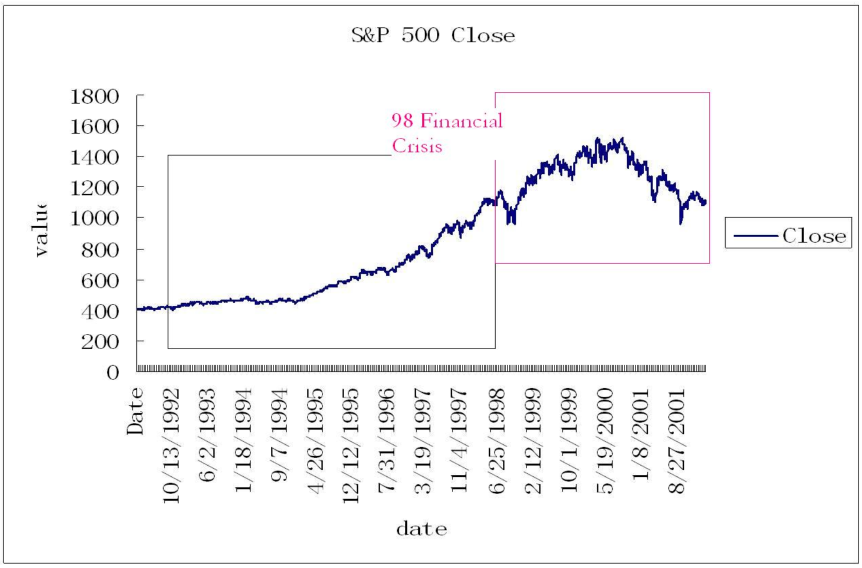


Figure 4.9. SP 500 Related to Possible News Event

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 a recursive estimator for state space models.

A state space model is like this:

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

where v_t and w_t are independent normal variables. $v_t \sim N(0, R_t)$ and $w_t \sim N(0, Q_t)$. In the context of our model, we have $z_t = log(y_t) = y_t^*$, $H_t = 1$, $x_t = h_t$, $F_t = \phi$ and $\mu * (1 - \phi)$.

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

Predict Phase:

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

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:

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

where $\tilde{y}_t$ is the measurement residual, and S_t is residual covariance. K_t is Kalman gain. $\hat{x}_{t|t}$ is updated value for this current time, and $P_{t|t}$ is the updated estimation covariance.

The key element is the Kalman gain K_t. K_t 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 hidden process information conditional on the information from time $t = 1$ to $t = n$. In the forward sampling pass, residuals $\tilde{y}_t$, covariance for residuals S_t, and kalman gain K_t are stored. Setting $r_n = 0, N_n = 0$, and writing $D_t = S_t^{-1} + K_tN_tK_t$ and $n_t = S_t^{-1}\tilde{y}_t - K_tr_t$, we run backward from $t = n, n-1, ..., 1$.

$$\begin{aligned}&C_t = R_t - R_tD_tR_t,\\&\kappa_t \sim N(0, C_t),\\&r_{t-1} = H_tS_t^{-1}\tilde{y}_t + L_t^{'}r_t - V_t^{'}C_t^{-1}\kappa_t\\&V_t = R_t(D_tH_t - K_t^{'}N_tB_t),\\&N_{t-1} = H_t^{'}S_t^{-1}H_t + L_t^{\prime}N_tL_t + V_t^{\prime}C_t^{-1}V_t\end{aligned}\quad .$$

Where $L_t = B_t - K_tH_t$. Then $z_t - R_tn_t - \kappa_t$ is a draw from the multivariate normal posterior distribution $p(H_tx_t|\mathbf{z}, \theta)$.

Bibliography

[1] 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.

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

[3] P. DE JONG and N. SHEPHARD. The simulation smoother for time series models. *Biometrika*, 82(2):339–350, 1995.

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

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

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

[7] 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 2003-01, Brown University, Department of Economics, 01 2003.

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

[9] 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. Econometric Society 2004 Latin American Meetings 198, Econometric Society, Aug. 2004.

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

[11] 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.

[12] S. Kim and N. Shephard. Stochastic volatility: likelihood inference and comparison with arch models. Economics Papers 3, Economics Group, Nuffield College, University of Oxford, Nov. 1994.

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

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

[15] R. M. Neal. Slice sampling. *Annal of Statistics*, 31(3):705–767, 2003.

[16] D. B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–70, March 1991.

[17] M. K. Pitt and N. Shephard. Filtering via simulation: auxiliary particle filters. Economics Papers 1997-W13, Economics Group, Nuffield College, University of Oxford, 6 1999.

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

[19] N. Shephard. Mstochastic volatility. 2005.

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

[21] J. Vermaak, C. Andrieu, A. Doucet, and S. J. Godsill. Reversible jump markov chain monte carlo strategies for bayesian model selection in autoregressive processes. *Journal of Time Series Analysis*, 25(6):785–809, November 2004.

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