A stochastic volatility model with leverage effect and regime switching

Hong Jiang
University at Albany, State University of New York, jhconnie@hotmail.com

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Economics Commons, Finance and Financial Management Commons, and the Statistics and Probability Commons

Recommended Citation
https://scholarsarchive.library.albany.edu/legacy-etd/1152

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive. Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
A Stochastic Volatility Model
With Leverage Effect and Regime Switching

By

Hong Jiang

A Dissertation
Submitted to the University at Albany, State University of New York
in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy

College of Arts & Sciences
Department of Economics
2014
Abstract

Modeling the volatility of asset returns is a very important study in financial economics. Among the time-varying volatility models, the Stochastic Volatility (SV) models are argued to have advantages over the autoregressive conditional heteroskedasticity (ARCH) models. The purpose of this article is to put forward a generalized and flexible Stochastic Volatility model, the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model), which could capture the complex features of financial time series to the most extent.

Even though the estimation to this SVLR model is a hard problem, a Bayesian Markov Chain Monte Carlo (MCMC) estimation method based on the Gibbs sampling algorithm is developed in the article. The SVLR model is applied to the S&P 500 index data and the relevant evolution of volatility process is studied.

In order to evaluate the likelihood of the SVLR model, a specific particle filter for the SVLR model is proposed, the diagnostic tests and the model comparison between the SVLR model and other existing time-varying volatility models under the Bayesian framework are conducted. The proposed SVLR model proves to outperform existing models in terms of its flexibility of describing the features of stock volatilities and better fitting the financial data.
The out-of-sample forecasting performance of SVLR model can be proved good enough, which suggests the promising application of the SVLR model in financial risk management areas. A new perspective to study the connection between stock market volatility and business cycle is proposed through combing the SVLR model with the regime switching autoregressive industrial production model in this article either.
Contents

1. Introduction ........................................................................................................................................ 1
   1.1 Characteristics of Volatility ........................................................................................................ 1
   1.2 ARCH Models .............................................................................................................................. 3
   1.3 Stochastic Volatility Models ........................................................................................................ 6
   1.4 Contributions ............................................................................................................................... 7
   1.5 Dissertation Organization ............................................................................................................ 9

2. Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR) .... 11
   2.1 The Basic Stochastic Volatility (SV) Model .................................................................................. 11
       2.1.1 Continuous-time Stochastic Volatility Models .................................................................. 11
       2.1.2 Discrete-time Stochastic Volatility Model ....................................................................... 12
   2.2 The Stochastic Volatility Model with Leverage Effect (SVL Model) .......................................... 13
   2.3 The Stochastic Volatility Model with Regime Switching (SVR Model) ................................. 15
   2.4 The Stochastic Volatility Model with Leverage Effect and Regime Switching .... 18
       (SVLR Model) .............................................................................................................................. 18

3. Estimation for SVLR Model Through Markov Chain Monte Carlo Methods .......... 21
   3.1 The Classical Versus Bayesian Inference Approaches ................................................................... 22
   3.2 Markov Chain Monte Carlo (MCMC) Method ............................................................................... 25
       3.2.1 Introduction of Markov Chain Monte Carlo (MCMC) Method ....................................... 25
       3.2.2 Gibbs and Metropolis-Hastings (M-H) sampling algorithms ............................................ 27
   3.3 Specific Sampling Algorithms to the SVLR Model ..................................................................... 29

4. Monte Carlo Simulation Studies for SVLR Model ................................................................. 43
   4.1 Performance of Bayesian Parameter Estimation ........................................................................... 43
4.2 Performance of Bayesian Volatility Estimation ................................................... 47

5. The Empirical Application of SVLR Model to Stock Returns ........................................ 49
   5.1 Data Description ........................................................................................................ 49
   5.2 Estimation Results of SVLR Model to Actual Data .................................................. 51
      5.2.1 Convergence of Gibbs Sampler ........................................................................... 51
      5.2.2 Estimation Results .............................................................................................. 53
   5.3 Estimation Results From the Reweighting MCMC Algorithm Corrected For Approximation .......................................................... 59

6. Filtering, Diagnostics and Model Comparison .............................................................. 63
   6.1 Filtering ...................................................................................................................... 63
      6.1.1 Particle Filter for Filtering in SVLR model ........................................................ 63
      6.1.2 One-step-ahead Prediction Density ................................................................. 67
   6.2 Diagnostics of SVLR Model for S&P 500 Index Weekly Return Data ..................... 68
   6.3 Model Comparison .................................................................................................... 73
      6.3.1 Likelihood Evaluation for the SVLR Model ...................................................... 75
      6.3.2 Estimation of Posterior Densities ..................................................................... 76
      6.3.3 Comparison between the SVLR model and the existing SV models ............ 79

7. Bayesian Density Forecasting of SVLR Model ............................................................. 82
   7.1 The Bayesian Prediction Density for the SVLR Model ............................................. 83
      7.1.1 The Bayesian Prediction Density for the Latent-Structured Model ................ 84
      7.1.2 The Bayesian Density Forecasting of the SVLR Model ................................ 85
   7.2 The Evaluation of the Bayesian Density Forecasting of SVLR Model .................... 87
      7.2.1 Loss Functions and Density Forecasts ............................................................. 88
7.2.2 The Probability Integral Transform (PIT)................................. 90

7.2.3 The Evaluation of Bayesian Density Forecasts of SVLR Model .......... 91

8. Stock Market Volatility and Business Cycle ........................................... 96

8.1 Industrial Production Model .................................................................. 96

8.2 Stock Market Volatility and Business Cycles........................................ 101

8.2.1 Bivariate Framework – Model B .................................................. 103

8.2.2 Bivariate Framework – Model C .................................................. 106

8.3 Empirical Studies .................................................................................. 109

Appendix .................................................................................................... 118

References ............................................................................................... 120
1. Introduction

The volatility of an asset return means the conditional standard deviation of the underlying asset return, which characterizes the uncertainty that the market assigns to the price of an asset. It has many important financial applications. It is a critical factor in option pricing.\(^1\) It can play an important role in asset allocation under the mean-variance framework. Volatility modeling could provide a simple approach to measuring value at risk (VaR) in risk management. Modeling the volatility of a time series can also improve the efficiency of parameter estimation and the accuracy of interval forecasting. The volatility index, like the VIX Index, could even be compiled and traded in the stock market.

1.1 Characteristics of Volatility

An important and special feature of stock volatility is that it cannot be directly observed. When there is only one observation in a trading day in the old time, the daily volatility is not observable from the return data. When intra-day data of the stock, like 10-minute returns, are available, we can estimate the daily volatility. However, the high frequency intra-day returns data have the information about the intra-day volatility but only limited information about the overnight volatility of returns data, the deviation

\(^1\) The well-known Black-Scholes option pricing formula states that the price of a European call option is 
\[ c_t = P_t \Phi(x) - K e^{-rT} \Phi(x - \sigma_t \sqrt{T}), \text{ and } x = \frac{\ln(P_t/K) + rt}{\sigma_t \sqrt{T}}, \] 
where \(P_t\) is the current price of the underlying stock, \(K\) is the strike price, \(T\) is the time to expiration, \(r\) is the continuously compounded risk-free interest rate, \(\sigma_t\) is the annualized conditional standard deviation of the log return of the specified stock, and \(\Phi(x)\) is the c.d.f. of the standard normal distribution.
between trading days. If one accepts that the option prices are governed by the Black-Scholes formula, one can use the option price to deduce the implied volatility of the underlying stocks. But the critique for this way is that some assumptions for the formula might not hold in reality.² The VIX index is an implied volatility. One can also resort to econometric models to estimate the volatilities of asset returns. The models for estimating the time-varying volatility of an asset return are called as conditional heteroscedastic models or volatility models.

Although volatility is not directly observable, it has some characteristics that are commonly seen in asset returns. These properties play an important role in the development of volatility models. First, volatility clusters exist. That is, volatility may be high for certain time periods and low for other periods. High volatilities are highly possibly followed by high volatilities and low volatilities are highly possibly followed by low volatilities. Second, volatility process is often stationary. That means volatility does not diverge to infinity. Third, there exists “leverage effect”, i.e. stock volatility reacts differently to a price increase or a price drop. The leverage effect suggests that stock price movements are negatively correlated with volatility. Since a company’s leverage ratio is defined as “debt over equity”, the falling stock prices imply an increased leverage of firms, more uncertainty, and hence volatility, will rise. Some volatility models are proposed specifically to correct the weakness of the existing ones for their inability to capture the characteristics mentioned here.

² For example, one of the assumptions for the Black-Scholes option pricing formula that the price of the underlying asset should follow a Geometric Brownian Motion might not hold in reality.
Conditional heteroscedastic models or volatility models can be classified into two
general categories. The models in the first category use an exact function form to govern
the evolution of volatilities, whereas those in the second category use a stochastic process
to describe the volatilities. The autoregressive conditional heteroscedastic (ARCH) model
(Engle 1982) and many generalizations of it, like the generalized GARCH (Bollerslev
1986) and the exponential GARCH (or EGARCH, Nelson 1991), belong to the first
category. For surveys of the models in ARCH family, one can refer to Bollerslev, Engle
and Nelson (1994). The stochastic volatility (SV) model and many of its generalization
are in the second category.

1.2 ARCH Models

A first time-varying volatility model was proposed by Engel (1982), which was called
an autoregressive conditional heteroskedasticity (ARCH) model. Let $y_t$ be the log return
of an asset at time $t$, defined as $y_t = \log(X_t/X_{t-1})$, where $X_t$ is the observed market
asset price. In the ARCH model, the mean equation and variance equation for the asset
return are specified as

\[
\text{Mean equation: } y_t = \sigma_t \epsilon_t \tag{1.1}
\]

\[
\text{Variance equation: } \sigma_t^2 = \mu + \phi y_{t-1}^2 \tag{1.2}
\]

where $\sigma_t$ is the volatility of the asset returns, which is a measure of risk on returns. Each
observed asset return $y_t$ has a standard deviation $\sigma_t$. The disturbance $\epsilon_t$ is Gaussian, i.e.
\( \varepsilon_t \sim iid N(0,1) \). The volatility models are concerned with the evolution of \( \sigma_t^2 \). The way under which \( \sigma_t^2 \) evolves over time distinguishes one volatility model from another. In the ARCH models, volatility is an exact function of the squares of past returns and therefore is an observed variable.

ARCH model is extended by Bollerslev (1986) to become generalized autoregressive conditional heteroscedasticity (GARCH) model:

\[
\begin{align*}
    y_t &= \sigma_t \varepsilon_t \\
    \sigma_t^2 &= \mu + \phi y_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (1.3)
\end{align*}
\]

where the volatility is a deterministic function of the squares of past return as well as of past volatility. The existence of GARCH model solve for the problem when higher-ordered ARCH models are needed.

Both ARCH and GARCH models use the simple function forms to describe the volatility evolution. It could be also seen from the structures of ARCH and GARCH models that large past shocks to asset returns tend to be followed by another large shock. This feature is similar to the volatility clustering observed for asset returns. The tail distributions of both ARCH and GARCH processes are heavier than that of a normal distribution, which conforms to the empirical findings that “outliers” appear more often in asset returns. Given the normality assumption in the mean equation, the likelihood functions for ARCH or GARCH models are easy to obtain, so the Maximum Likelihood
parameter estimations are feasible. More extended ARCH models are developed to capture more characteristics of stock volatilities. For example, the exponential GARCH (or EGARCH) model by Nelson (1991) was developed to the asymmetric reaction of volatilities to positive and negative returns. Another ARCH model usually used to handle the leverage effects is the threshold GARCH (or TGARCH) model by Glosten, Jagannathan, and Runkle (1993).

Although the ARCH type models have lots of good properties mentioned above, the models also have their weakness. The ARCH models do not provide any insights for understanding the source of variations of a financial time series. The time-changing volatility process in ARCH models is just driven by the price change. The ARCH model merely provides a mechanical way to describe the evolution of stock volatilities, but it has no indication about what causes such behavior to occur. Actually, in the literature of volatility models, it was argued that the volatility process should be driven by economic forces like in the family of SV models (Taylar 1986) rather than just the movement of prices like in the family of ARCH models. Moreover, SV models could give a discrete time approximation for the diffusion processes used in option pricing (Hull and White 1987a,b). The introduction of the innovation in the conditional variance equation could increase the flexibility of the SV models in describing the evolution of volatility. Comprehensive reviews of the SV models can be found in the works of Taylor (1994) and Shephard (1996).

---

3 In the works of Hull and White (1987a,b) etc, the Black-Scholes option pricing formula was generalized to incorporate stochastic volatility.
1.3 Stochastic Volatility Models

Stochastic Volatility (SV) models provide a natural alternative to the ARCH family models. The first SV model was proposed by Taylor (1986), which makes volatility dependent on a latent variable $h_t$ called news via the relation $\log \sigma_t^2 = h_t$. Compared to ARCH type models where volatility could be observed, in the SV models, the volatility is unobserved. The basic SV model specifies a log-normal autoregressive process for the conditional variance of asset return and introduces an innovation in the variance equation. In SV models, the stock volatility process is driven by economic forces not just price changes like in ARCH type models, which increases flexibility on ARCH and allows for the further studies about the real economic cause of time-varying volatility. Another important motivation of SV models is their close connections to the continuous-time models in the asset-pricing literature in the modern financial theory.

Unlike the ARCH type models, the likelihood functions of SV models are much harder to evaluate, which lead to the difficulty of Maximum Likelihood estimation. The reason is that the likelihood functions of SV models are high-dimensional integrals and the dimensions will grow tremendously as the data sample size increases.\(^4\) The estimation twenty years ago used to be realized via Generalized Method of Moments (GMM) or Quasi-maximum Likelihood (QML), but both methods proved to be inefficient. The development of Bayesian method makes the Markov Chain Monte Carlo (MCMC) method an efficient way to analyze the SV model, like parameter estimation, filtering and prediction etc. Moreover, the more powerful computation capabilities of computers

\(^4\) Chapter 3 in the future will explain it in more details.
nowadays boost the rapid growth of various SV models. To see the introduction of Bayesian methods for parameter estimation of SV models, one can refer to Jacquier, Polson and Rossi (1994) and Kim, Shephard and Chib (1998).

1.4 Contributions

The basic SV model is simple but restrictive to capture some important features of financial time series. The purpose of this article is to put forward a most generalized and flexible stochastic volatility model, the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model), providing a new way to describe the volatility process in the SV models. In the SVLR model, the asymmetric feature, i.e. leverage effect, and switching regimes in the volatility process of financial data are both incorporated in order to better capturing the complex structure of financial data. Hamilton and Lin (1995) propose a generalized ARCH type model, the Regime Switching ARCH Model with Leverage Effect, to illustrate the leverage effect and the structure change in volatility process simultaneously, but there has not existed a corresponding generalized Stochastic Volatility model in the literature that can explain both of the important facts for asset returns, the “leverage effect” and the structure change in volatility. This article fills in the gap.

The Bayesian estimation method, Markov Chain Monte Carlo (MCMC) method, is adopted to analyze the SVLR model in this dissertation. Specifically, the Gibbs sampling algorithm is exploited. In order for the inference of the SVLR model, besides the estimates for the regular parameters, the estimates for the latent volatilities (one for each
period) and for the states (one for each period) are obtained. Thus the number of variables to be estimated is enormous. The situation contrasts traditional classical estimation methods, in which the number of parameters estimated is small and does not grow as the sample size increases. How to design a good sampler for the SVLR model that can converge fast is very challenging. Deriving the posterior conditional density for each variable and sampling from the conditional density are both challenging jobs. An efficient Bayesian estimation method is proposed and proved in this dissertation.

In order to complete the methodological development for the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR) model, a particle filter for the SVLR model, is initiated and filtering features of the latent variables in the model are achieved. The Diagnostic test for the SVLR model and the Model Comparison between it and other existing models are facilitated by the success of the particle filter. The SVLR model is applied in the actual financial data, the S&P 500 index data. It is identified that the SVLR model outperforms other existing volatility models in terms of its flexibility of capturing the important asymmetric features, i.e. leverage effect, in volatility process, explaining the real cause to the high persistence of volatility, and better in-sample data fit.

One of the important goals of the econometric analysis to the SVLR model is to predict the future values of financial data of interest. The forecasting work under the Bayesian framework in this dissertation is to conduct the density forecast, in contrast to

---

5 Particle filters or Sequential Monte Carlo (SMC) methods are a set of on-line posterior density estimation algorithms that estimate the posterior density of the state-space by directly implementing the Bayesian recursion. These filtering methods make no restrictive assumption about the dynamics of the state-space or the density function. The employment of the particle filter in the SVLR model is illustrated in Chapter 6.
the point forecast or the interval forecast. The Bayesian density forecasting takes into account the complete uncertainty related with a prediction. It is verified that the SVLR model could be applied in the financial risk management area, for example, to provide the density forecasting of portfolio values or calculating the Value-at-Risk (VaR).6

An important application of the SVLR model is to provide a new perspective to study the connection between the stock market volatility and the business cycle. This dissertation builds up a bi-variates framework in which the SVLR model for the stock market and the industrial production model for the real economy are investigated jointly. The stock market volatility process and the real output process are confirmed to be driven by related economic forces. From this point, the bi-variates framework could help improve identifying and forecasting business cycles.

1.5 Dissertation Organization

The main body of the dissertation is organized into eight chapters to discuss the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model). The dissertation starts in Chapter 2 with specifying the model form of the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model). Chapter 3 explains how to exploit the Markov Chain Monte Carlo (MCMC) sampling methods to do parameter estimation for the Stochastic Volatility Model with Leverage Effect and

---

6 In financial risk management, VaR is a widely used risk measure of the risk of loss on a specific portfolio of financial assets. For a given portfolio, probability and time horizon, VaR is defined as a threshold value such that the probability that the loss on the portfolio over the given time horizon exceeds this value is the given probability level.
Regime Switching (SVLR model). Chapter 4 does a Monte Carlo simulation study to verify that the algorithm developed in Chapter 3 is efficient to do parameter estimation. Chapter 5 applies the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) to the real return data, the S&P 500 index data. The estimation results are illustrated and analyzed. Chapter 6 firstly develops a particle filter to get the filtering inference for the latent variables. Then the likelihood evaluation and model failure diagnostics for the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) are implemented. The comparisons between the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) and other time-varying volatility models are performed. Chapter 7 introduces the method to evaluate the Bayesian density forecasting and studies the performance of the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) on Bayesian density forecasting. Chapter 8 adopts the SVLR model to build up a framework to study the connection between stock market volatility and business cycles.
2. Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR)

In the literature, there is evidence that Stochastic Volatility (SV) models provide increased flexibility over the ARCH type models, e.g. Geweke (1994b) and Fridman and Harris (1998). In SV models, it is economic forces rather than the movement of prices that drive the time-varying volatility process. Moreover, stochastic volatility models give a discrete time approximation to the diffusion processes used in option pricing. This chapter will start with introducing the background and the specific form of the basic Stochastic Volatility (SV) models. In order to get the generalized and flexible SV Model, the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model), the basic SV model is extended in two natural ways: (1) with a “leverage effect”, in which the innovations in the mean equation and the variance equation of return are correlated and (2) with a regime-switching scale of the volatility.

2.1 The Basic Stochastic Volatility (SV) Model

2.1.1 Continuous-time Stochastic Volatility Models

In financial theoretical work, especially with derivatives, it is convenient to formulate the models in continuous time. An example is the geometric diffusion models used by Hull and White (1987). There work is to generalize the Black – Scholes option-pricing formula to allow for stochastic volatility, i.e. the asset price and the volatility each follow a diffusion process. In their model, stock prices follow a diffusion process,
\[
\frac{dP_t}{P_t} = \alpha dt + \sigma(t)dW_1(t),
\]  
(2.1)

and the logarithm of \(\sigma(t)\) follows a diffusion process given by the Ornstein – Uhlenbeck (O-U) process,

\[
d(ln\sigma_t) = \lambda(\xi - ln\sigma_t)dt + \gamma dW_2(t),
\]  
(2.2)

where \(P_t\) is stock price, \(\sigma(t)\) is the instantaneous variance of \(P_t\), \(W_1(t)\) and \(W_2(t)\) are two Wiener processes, \(\alpha, \lambda, \xi, \gamma\) are the constant parameters. Several authors specify the different diffusion processes and give the correspondent numerical option results, e.g. Stein (1991), Heston (1993) and Johnson and Shanno (1987).

### 2.1.2 Discrete-time Stochastic Volatility Model

For empirical work, it is convenient to use discrete-time data. Parameter estimates can be obtained for discrete-time approximations of the popular continuous-time models. The basic Stochastic Volatility model in this dissertation is the simplest discrete time approximation to the continuous time stochastic volatility models found in Taylor (1986). It could be specified as

\[
y_t = \exp(h_t/2) \epsilon_t \\
h_{t+1} = \mu + \phi h_t + \eta_t, \ t = 1, \ldots, T,
\]  
(2.3)
where $y_t$ is the log return of assets (i.e. $\log P_t - \log P_{t-1}$), $h_t$ is the unobserved log volatility (i.e. $\log \sigma_t^2$). The log volatility process is assumed to follow a stationary AR (1) process through the restriction of $|\phi| < 1$. $\varepsilon_t$ and $\eta_t$ are uncorrelated normally distributed shocks with both means of 0 and variances of 1 and $\sigma^2$ respectively. The parameter $\mu$ could be thought as the mean level of the log volatility, $\phi$ as the persistence and $\sigma$ as the volatility of the log volatility.

The basic SV model is simple but have its own limitation. It could not explain the leverage effect in volatility, an important and well-documented empirical features for financial time series (Black 1976, Christie 1982, Engle and Ng 1993). It also could not explain the structure change in volatility. The following sections in this dissertation will extend this basic Stochastic Volatility model in these directions.

2.2 The Stochastic Volatility Model with Leverage Effect (SVL Model)

The negative relationship between volatility and price or return is called Leverage Effect. The mechanism behind this negative relationship is that when bad news comes, the firm’s stock price will decrease so the debt-to-equity ratio (i.e. financial leverage) will increase, which makes the firm riskier then the uncertainty (i.e. volatility) increases. Black (1976) and Christie (1982) have found the empirical evidence of this leverage effect, i.e. volatilities tend to rise when stock price falls but decrease when stock price rise. Christie (1982) gives a theoretical explanation of leverage effect under a Modigliani/Miller framework. Hull and White (1987) also argue that the valuation of options considering stochastic volatility process would be biased if neglecting the
negative correlation between volatility and stock price. So, one of the natural ways to extend the basic Stochastic Volatility model is to incorporate Leverage Effect.

For the ARCH type models, the leverage effect is specified in the EGARCH model of Nelson (1991) and the threshold GARCH model of Glosten, Jagannathan, and Runkle (1993). In the literature of the SV models, Harvey and Shephard (1996) propose a SV model with leverage effect and use quasi-maximum likelihood method to fit the stock data. With the same motive, Jacquier, Polson and Rossi (2004) generalize the basic SV model by incorporating an asymmetric feature to catch the leverage effect and propose the single move MCMC methods for estimation. Ompri, Chib, Shephard and Nakajima (2007) extend the multiple move MCMC methods in Kim, Shephard and Chib (1998) for the SV model with leverage effect.

The specification of the Stochastic Volatility Model with Leverage Effect (SVL) in this dissertation would be in the way of Harvey and Shephard (1996):

\[
y_t = \exp(h_t/2) \varepsilon_t
\]

\[
h_{t+1} = \mu + \phi h_t + \eta_t, \ t = 1, \ldots, T,
\]

\[
\begin{pmatrix}
\varepsilon_t \\
\eta_t
\end{pmatrix}
\sim \text{i.i.d.} \mathcal{N}_2(0, \Sigma), \quad \Sigma = \begin{pmatrix}
1 & \rho \sigma \\
\rho \sigma & \sigma^2
\end{pmatrix}, \quad (2.4)
\]

where \(\mathcal{N}_2(0, \Sigma)\) is the bivariate normal distribution with mean vector \(\mathbf{0}\) and covariance matrix \(\Sigma\). This model introduces correlated errors and parameter \(\rho\) measures the correlation between \(\varepsilon_t\) and \(\eta_t\). If the correlation \(\rho\) is negative, then a negative innovation
in the mean equation, \( \varepsilon_t \), will be associated with the higher volatility in the log volatility process of next period time. This asymmetry is exactly the leverage effect.

It is worth noting that the specification way for leverage effect in Jacquier et al. (2004) is different where the log volatility process is specified as \( h_t = \mu + \phi h_{t-1} + \eta_t \), instead of the one in equation (2.6). The problem of this specification is that the model is not consistent with the efficient market hypothesis because \( y_t \) is not a martingale difference process any more, whereas our specification will remain the martingale difference property for \( y_t \). In addition, the interpretation of leverage effect is not clear under this framework. On the contrary, the interpretation of leverage effect is straightforward and clear in our specification. Yu (2005) gives more detailed discussion of these issues from both theoretical and empirical aspects.

### 2.3 The Stochastic Volatility Model with Regime Switching (SVR Model)

The high persistence in volatility is a common finding in the ARCH and SV literatures, e.g. Chou (1988), French, Schwert, and Stambaugh (1987), Poon and Taylor (1992), and So, Lam, and Li (1997). That means volatility may be high for certain time periods and low for other periods. We could show the high persistence in volatility using a graph of the square S&P 500 index return in Figure 2.1.

However, Lamoureux and Lastrapes (1990) point out that the high persistence of the volatility may be overestimated since the possible structural change in the volatility process had not been taken into account. Hamilton and Susmel (1994) propose the
Markov Switching ARCH model that introduces the regime switching idea of Hamilton (1988, 1989) in the ARCH model to allow for the structural shift in the scale of the conditional variance of returns. Its application to the New York Stock Exchange data verify that most of the persistence in volatility is accounted for by the persistence in each regime. So, Lam and L (1998) firstly combine the SV model with the Markov switching model by Hamilton (1989) to accommodate the shift in the mean level of the log-volatility process. They apply the Stochastic Volatility model with regime switching, i.e. SVR model, to S&P 500 index data and find that the estimate of persistence in volatility drops significantly. Kalimipalli and Susmel (2004) apply the SVR model to explain the behavior of short-term interest rates and obtain the same results that the persistence in volatility is substantially reduced by the introduction of regime switching.

**Figure 2.1 Squared S&P 500 Index Monthly Return (%) in 01/1985 - 12/2011**
The Stochastic Volatility Model with Regime Switching, i.e. SVR model, could be expressed as:

\[ y_t = \exp(h_t/2) \varepsilon_t \]
\[ h_{t+1} = \mu s_{t+1} + \phi h_t + \eta_t, \quad t = 1, \ldots, T, \]

where \( \{h_t\} \) are still unobserved log-volatilities, \(|\phi| < 1\) still assure that the log volatility process is stationary, and \( \left( \begin{array}{c} \varepsilon_t \\ \eta_t \end{array} \right) \sim i. i. d. \mathcal{N}_2 \left( \begin{array}{c} 0 \\ 0 \end{array}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2 \end{pmatrix} \right) \). Here, \( s_t \), another latent variable, is the state variable that determines which regime the log volatility process will belong to. For different regime, the mean level for the log volatility process will be different. The stochastic process of \( s_t \) could be specified as a hidden Markov process (the details are in the following section). This kind of extension for the basic SV model is helpful for studying the real persistence in the log volatility process and the effect of big economic event on the data series.

Shibata and Watanabe (2005) provide the particle filter for the Markov switching SV model and do the model comparison and diagnostics for it. Hwang, Satchell and Pereira (2007) present a SV model with regime-dependent volatility levels, persistence level and volatilities of volatility and use the quasi-maximum likelihood method to do estimations.
2.4 The Stochastic Volatility Model with Leverage Effect and Regime Switching
(SVLR Model)

In this dissertation, we will put forward the most generalized Stochastic Volatility model, the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model). This model makes use of the extensions of the basic Stochastic Volatility model in section 2.2 and 2.3 simultaneously, i.e. the innovations in the mean equation and log volatility process are correlated and the log volatility process has two different regimes, the high- and the low- volatility regimes.

Suppose that we have a financial return series \( Y = (y_1, \cdots, y_T) \), and define \( S_t \) as a latent state variable that takes 1 in high-volatility regime and 0 in low-volatility regime, then the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) could represented by

\[
y_t = \exp(h_t/2) \varepsilon_t \\
h_{t+1} = \mu_{S_{t+1}} + \phi h_t + \eta_t, \quad t = 1, \cdots, T,
\]

where \( \{h_t\} \) are unobserved log-volatilities, \( |\phi| < 1 \),

\[
\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim \text{i.i.d. } \mathcal{N}_2(0, \Sigma), \quad \Sigma = \begin{pmatrix} 1 & \rho \sigma \\ \rho \sigma & \sigma^2 \end{pmatrix},
\]
and \( N_2(0, \Sigma) \) is the bivariate normal distribution with mean vector \( \mathbf{0} \) and covariance matrix \( \Sigma \). The parameter \( \rho \) measures the correlation between \( \varepsilon_t \) and \( \eta_t \) and when negative, it represents the leverage effect of in the model, i.e. the drop in asset price is followed by the increase in volatility whereas the rise in asset price is followed by the decrease in volatility. The mean level of log-volatility process, i.e. \( \mu_{S_t+1} \), will be determined by the value of the unobserved state variable \( S_{t+1} \), representing which regime the log volatility process belongs to. We can write the mean level of the log volatility process as

\[
\mu_{S_t} = \begin{cases} 
\mu_0 & \text{if } S_t = 0, \\
\mu_0 + \mu_1 & \text{if } S_t = 1, \text{ and } \mu_1 > 0.
\end{cases}
\]

Then the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) could be rewritten as

\[
y_t = \exp(h_t/2) \varepsilon_t \\
h_{t+1} = \mu_0 + \mu_1 S_{t+1} + \phi h_t + \eta_t, \quad t = 1, \cdots, T, \text{ and } \mu_1 > 0. \quad (2.7)
\]

where \( \{h_t\} \) are unobserved log-volatilities, \( |\phi| < 1 \),

\[
\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim i.i.d. N_2(0, \Sigma), \quad \Sigma = \begin{pmatrix} 1 & \rho \sigma \\ \rho \sigma & \sigma^2 \end{pmatrix}.
\]

The stochastic process of \( \{S_t\} \) is determined by a two-state first order Markov process whose transition probability matrix is
where \( P(s_t = 0 | s_{t-1} = 0) = q \) and \( P(s_t = 1 | s_{t-1} = 1) = p \). The transition probability parameter \( q \) and \( p \) represents probability of remaining in the low-volatility regime and high-volatility regime respectively.

The structure of the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) could capture the asymmetric response of volatility to price/return, the usual changing behavior of volatility due to economic forces, and the sudden discrete shift in volatility due to sudden abnormal events. More generalized models could be obtained through extending this model straightforwardly. For example, the number of states could be more than two like in So, Lam and Li (1998) and other parameters may also shift like in Hwang, Satchell and Valls Pereira (2007) that the SV model being with regime-dependent volatility levels, persistence level and volatilities of volatility.

For the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) specified in equations (2.7), the parameters of \((\phi, \sigma, \rho, \mu_0, \mu_1, p, q)\) as well as the latent variables \(\{h_t\}\) and \(\{S_t\}\), where \(t = 1, \ldots, T\), will need to be estimated.
3. Estimation for SVLR Model Through Markov Chain Monte Carlo Methods

In this chapter, we will develop the Gibbs sampler of MCMC methods to estimate the parameters and gain the inference for the Stochastic Volatility Model with Leverage Effect and Regime Switching. Specifically, the parameter vector \( \theta = (\phi, \sigma, \rho, \mu_0, \mu_1, p, q)' \), the log-volatilities vector \( H = (h_1, \ldots, h_T)' \) and the regime states vector \( S = (S_1, \ldots, S_T)' \) will need to be estimated. Different from the ARCH type models, given the parameters \( \theta = (\phi, \sigma, \rho, \mu_0, \mu_1, p, q)' \), the likelihood of the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) is the density of the return data \( Y = (y_1, \ldots, y_T) \):

\[
L(\theta) = f(Y|\theta) = \iint f(Y, H, S|\theta) dH dS,
\]

a very high-dimensional integral. This likelihood is intractable and not easy to compute. Here the Maximum Likelihood Estimation (MLE) could not work. Actually, all the Stochastic Volatility (SV) models have the similar challenges. In order to do the estimations for SV models, Taylor (1982, 1986, 1994), Andersen (1994) and Andersen and Sørensen (1996) perform the estimation through the Generalized Method of Moments (GMM). As is well known, the GMM method is usually be inefficient relative to a likelihood-based method of inference. Harvey, Ruiz and Shephard (1994), Harvey and Shephard (1996) and Hwang, Satchell and Pereira (2007) use the Quasi-maximum Likelihood (QML) methods to do the estimations. They transform the mean equation of the Stochastic Volatility models into a linear but non-Gaussian equation, use a normal distribution to approximate the disturbances, and employ the Kalman Filter to maximize
The problem for the QML method is that the normal distribution is a poor approximation to the non-Gaussian distribution and the filtered volatilities estimates have high standard errors. The difficulty in estimation is the reason to limit the development and application of Stochastic Volatility (SV) models. Jacquier, Polson and Rossi (1994) advance a Bayesian method using Markov Chain and Monte Carlo (MCMC) simulation technique to do the estimation. Kim, Shephard and Chib (1998) develop the MCMC analysis of the Stochastic Volatility (SV) models. More and more research for Stochastic Volatility (SV) models resort to this method to obtain the Bayesian inference. This chapter starts with the introduction of the Bayesian inference method. Then the MCMC simulation method tremendously used in Bayesian analysis will be explained. Finally, the specific estimation algorithm for the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) will be illustrated in details.

3.1 The Classical Versus Bayesian Inference Approaches

In the literature, there are mainly two approaches to statistical inference. They are the classical approach based on maximum likelihood principle and the Bayesian approach that combines prior beliefs with data to obtain posterior distribution on which the statistical inference is based on.

A statistical model with a parameter vector \( \boldsymbol{\theta} \) specifies a joint distribution for a vector of observations \( Y, f(Y|\boldsymbol{\theta}) \). This joint density function tells us the probability of obtaining a particular observed dataset \( Y \) given \( \boldsymbol{\theta} \). In econometric practice, the series \( Y \) is just one realization of the data vector called sample data and we do not know the values
of the parameters $\theta$. In this case, the joint density $f (Y|\theta)$ is a function of $\theta$ given $Y$, and it is called the likelihood function, $L (\theta|Y)$, equal to $f (Y|\theta)$.

Under the classical framework, parameters $\theta$ are treated as unknown constants. In the Maximum Likelihood Estimation (MLE) method, we would choose the parameter estimates that maximize the probability of having generated the observed sample data, through maximizing the log of the likelihood function, i.e. the parameter estimates in MLE is $\hat{\theta}_{ML} = \text{Argmax} \ln L (\theta|Y)$. The standard error of the MLE estimates could be calculated by using the inverse of the negative of the second derivative of the log likelihood (Hessian) evaluated at $\hat{\theta}_{ML}$. The statistical inference of the model from MLE is from sample data and thus is objective. A more complete explanation of the Maximum Likelihood Estimation (MLE) method could be found in Cramer (1986), Judge (1982), David and MacKinnon (1993), and Harvey (1990).

Under the Bayesian framework, the parameter $\theta$ is regarded as random variables having probability distributions. These distributions summarize the information about the model’s parameters. The distribution employed by researchers before observing the data is called the prior distribution of the parameters, $f (\theta)$. After obtaining the data, researchers will revise the distribution of parameters by combing the prior distribution with the new information contained in the data and the posterior distribution of parameters is $p (\theta|Y)$. From Bayes theorem, we have $p (\theta|Y) \propto f (Y|\theta)f (\theta)$. Because of the equivalence of joint density $f (Y|\theta)$ and likelihood function $L (\theta|Y)$, it is customary to express the posterior distribution as $p (\theta|Y) \propto L (\theta|Y) f (\theta)$. The statistical inference
of the model will be from the posterior distribution that combines the likelihood function and the prior distribution. It is obvious that the Bayesian inference involves the researchers’ prior information about the model and thus is subjective. Different researchers could have different prior beliefs. One can take advantage of his subjective prior information in Bayesian inference. Or, if the prior information is very limited, a “flat” prior distribution will be adopted and is called the non-informative or diffuse prior. Reference for Bayesian statistics is available from Carlin and Louis (2000) and Carlin, Stern and Rubin (2003).

Historically, there are heated debates between the classical inference and the Bayesian inference, but there is no consensus which one is better. Both of the approaches prove to be useful are widely applied. As mentioned above, even the simplest Stochastic Volatility model has a very intractable and hard-to-compute likelihood function, so the Maximum Likelihood Estimation (MLE) could not work for SV models. In this dissertation, we will choose the Bayesian method to do estimation and get the inference to the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model).

Under the Bayesian framework, we are concerned with the posterior distribution of the parameters given the data for the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model), i.e. $p(\theta|Y)$, where $\theta = (\phi, \sigma, \rho, \mu_0, \mu_1, p, q)'$. Adopting the idea of the “Data Augmentation” by Tanner and Wong (1987), the parameter vector $\theta = (\phi, \sigma, \rho, \mu_0, \mu_1, p, q)'$ is augmented with the latent log-volatilities
vector $H = (h_1, \cdots, h_T)'$ and the regime states vector $S = (S_1, \cdots, S_T)'$. The joint posterior distribution of interest becomes $p(\theta, H, S \mid Y)$. The relevant analysis for the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model), including estimation, filtering, diagnostics and model comparison and so on, will all be based on this augmented joint posterior distribution. The virtue of this scheme is that it allows us to conduct the inference without having to calculate the likelihood function of the parameters.

3.2 Markov Chain Monte Carlo (MCMC) Method

As mentioned above, we want to get the Bayesian inference for the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) from the high dimensional joint posterior distribution $p(\theta, H, S \mid Y)$, where $Y$ represents the observed dataset, $\theta$ the parameter vector and $H$ and $S$ the latent log-volatilities and regime states vectors respectively. Markov Chain Monte Carlo (MCMC) method, which deals with the simulation of the high dimensional probability distributions, will provide us an efficient way to fulfill our objectives.

3.2.1 Introduction of Markov Chain Monte Carlo (MCMC) Method

Markov Chain Monte Carlo (MCMC) methods have proved enormously popular in Bayesian statistics, since these methods can obtain the inference to the posterior distributions through simulating from the distribution instead of calculating the distribution directly. Especially when the posterior distribution is complex, MCMC
methods could demonstrate its advantage and power in the Bayesian analysis framework. The advances in computing capabilities of computers and computational methods speed the development of MCMC method.

Markov Chain Monte Carlo (MCMC) method is a method to sample from a given probability distribution that is referred to as the target distribution through constructing a Markov Chain whose limiting invariant distribution is the target distribution. Once the Markov Chain has been constructed, a sample of (correlated) draws from the target distribution could be obtained by repeatedly sampling from the Markov Chain and retaining these values after the Markov Chain starts converging fast. For the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) in this dissertation, the MCMC method could be used to obtain the samples for the joint posterior distribution $p(\theta, H, S|Y)$ and then the inference to the model, like the posterior marginal density functions of parameters, could be obtained from the MCMC output.

The key to the Markov Chain Monte Carlo (MCMC) method is to create a Markov process whose stationary transition distribution is exactly the target distribution, the joint posterior distribution $p(\theta, H, S|Y)$ for the SVLR model, and run the simulation long enough to ensure the distribution of the sampled draws from the Markov process close to the stationary transition distribution, i.e. the target distribution. Gibbs or Metropolis-Hastings (M-H) sampling algorithms are two approaches to construct the Markov Chains with limiting distribution the target distribution.
3.2.2 Gibbs and Metropolis-Hastings (M-H) sampling algorithms

In order to illustrate each algorithm, we define the multivariate target distribution as $\pi (\boldsymbol{\psi})$ and the variables are grouped into $p$ blocks ($\boldsymbol{\psi}_1, \cdots, \boldsymbol{\psi}_p$). In the Gibbs sampler, each block is sampled according to the full conditional distribution of block $\boldsymbol{\psi}_k$ represented by $\pi (\boldsymbol{\psi}_k | \boldsymbol{\psi}_{\neg k})$, where $\boldsymbol{\psi}_{\neg k}$ denotes all the blocks excluding $\boldsymbol{\psi}_k$. The full conditional distribution is usually simple and could be obtained by Bayes Theorem, $\pi (\boldsymbol{\psi}_k | \boldsymbol{\psi}_{\neg k}) \propto \pi (\boldsymbol{\psi}_k, \boldsymbol{\psi}_{\neg k})$, the joint distribution of all blocks. Usually, the technique of Data Augmentation proposed by Tanner and Wong (1987) introduces latent or auxiliary variables into the model to simplify the derivation and sampling of the full conditional distributions. Here in order to explain the method of Gibbs sampler, we use a simple example with three blocks for the parameter vector, where the parameter vector is the generalized vector that could include latent or auxiliary variables. For example, the targeting joint posterior distribution is $f (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \boldsymbol{\psi}_3 | \boldsymbol{Y})$, where $\boldsymbol{Y}$ is the observed dataset. To sample from this joint posterior distribution, a Markov chain could be constructed through using the full conditional distributions for each parameter block. We assume that the three conditional distributions, $f (\boldsymbol{\psi}_1 | \boldsymbol{\psi}_2, \boldsymbol{\psi}_3, \boldsymbol{Y})$, $f (\boldsymbol{\psi}_2 | \boldsymbol{\psi}_1, \boldsymbol{\psi}_3, \boldsymbol{Y})$ and $f (\boldsymbol{\psi}_1 | \boldsymbol{\psi}_2, \boldsymbol{\psi}_3, \boldsymbol{Y})$ are known. Let $\boldsymbol{\psi}^0_1$, $\boldsymbol{\psi}^0_2$ and $\boldsymbol{\psi}^0_3$ be the arbitrary starting values of $\boldsymbol{\psi}_1$, $\boldsymbol{\psi}_2$ and $\boldsymbol{\psi}_3$. The Gibbs sampler proceeds in the way:

1. Draw a random sample from $f (\boldsymbol{\psi}_1 | \boldsymbol{\psi}_2^0, \boldsymbol{\psi}_3^0, \boldsymbol{Y})$ and denote this draw by $\boldsymbol{\psi}_1^1$.
2. Draw a random sample from $f (\boldsymbol{\psi}_2 | \boldsymbol{\psi}_1^1, \boldsymbol{\psi}_3^0, \boldsymbol{Y})$ and denote this draw by $\boldsymbol{\psi}_2^1$.
3. Draw a random sample from $f (\boldsymbol{\psi}_3 | \boldsymbol{\psi}_1^1, \boldsymbol{\psi}_2^1, \boldsymbol{Y})$ and denote this draw $\boldsymbol{\psi}_3^1$. 

27
This completes one-time Gibbs iteration and the parameter becomes $\Psi_1^1, \Psi_2^1$ and $\Psi_3^1$. Next, using the new parameters as starting values and repeating the above iteration, we can complete another-time Gibbs iteration and get the new parameters $\Psi_1^2, \Psi_2^2$ and $\Psi_3^2$. We can repeat these similar iterations many times, like $m$ times, and get a sequence of random draws,

$$\left(\Psi_1^1, \Psi_2^1, \Psi_3^1\right), \ldots, \left(\Psi_1^m, \Psi_2^m, \Psi_3^m\right).$$

Under some regularity conditions, it can be shown that, for a sufficient large $m$, $(\Psi_1^m, \Psi_2^m, \Psi_3^m)$ could be approximately regarded as the random draws form the joint distributions $f(\Psi_1, \Psi_2, \Psi_3 | Y)$ of the parameters. This regularity conditions are weak and require that for an arbitrary starting values $(\Psi_1^0, \Psi_2^0, \Psi_3^0)$, the Gibbs iterations have a chance to visit the full parameter space. In practice, we use a sufficient large $m$ and the first $n$ random draws are discarded and called the burn-in samples. The purpose of burn-in samples is to ensure that the rest random draws,

$$\left(\Psi_1^{n+1}, \Psi_2^{n+1}, \Psi_3^{n+1}\right), \ldots, \left(\Psi_1^m, \Psi_2^m, \Psi_3^m\right),$$

called Gibbs samples, would be close enough to the random samples from the joint distribution $f(\Psi_1, \Psi_2, \Psi_3 | Y)$. The Gibbs samples could be used to make inference. For example, the average mean of the samples for each parameter can be calculated as the estimate of the parameter and the samples could be used to estimate posterior marginal
density functions, called Rao-Blackwellisation. An elementary introduction to Gibbs sampler is available from Casella and Smith (1990).

For some blocks, the conditional distribution has no standardized form and not easily sampled directly. One can then use the Metropolis-Hastings (M-H) step whereby $\psi$ is sampled from a blanketing density $q(\psi, \psi')$, i.e. upper bounding density. $q(\psi, \psi')$ is also called a proposal density which provides a candidate value $\psi'$. The new draw $\psi'$ from $q(\psi, \psi')$ is accepted with probability

$$
\alpha(\psi, \psi') = \min \left[ \frac{\pi(\psi')q(\psi', \psi)}{\pi(\psi)q(\psi, \psi')}, 1 \right].
$$

Otherwise, the previous draw, i.e. $\psi$ is repeated, and one moves to the next block. The sampled draws could be also used in making inference. Compared to M-H algorithm, Gibbs sampler is more simple and straightforward and we will mainly apply this method for the SVLR model in the dissertation.

3.3 Specific Sampling Algorithms to the SVLR Model

For the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model), vector $\theta = (\phi, \sigma, \rho, \mu_0, \mu_1, p, q)'$ includes the regular parameters. Under the Bayesian framework, the log-volatilities vector $H = (h_1, \cdots, h_T)'$ and the regime states vector $S = (S_1, \cdots, S_T)'$ consist of the augmented parameters. A Gibbs sampling of MCMC methods will be used to estimate the regular parameters as well as the augmented
parameters through drawing the random samples from the following conditional posterior distributions: 

\[ f(\mu_0, \mu_1, \phi | \sigma, \rho, p, q, Y, H, S) \] , \[ f(\sigma, \rho | \mu_0, \mu_1, \phi, p, q, Y, H, S) \] , 

\[ f(p, q | \mu_0, \mu_1, \phi, \sigma, \rho, H, S) \] , \[ f(h_1, \cdots, h_T | \theta, Y, S) \] and \[ f(S_1, \cdots, S_T | \theta, Y, H) \] , where \( Y \) is the observed data. The Gibbs samples from the above conditional posterior distributions could be used to calculate the parameter estimates and the smoothed estimates of the latent log-volatilities and regime state variables. In this section, we will illustrate in details how to draw the samples from each conditional distribution.

1. Drawing \( \mu_0, \mu_1 \) and \( \phi \) from \( f(\mu_0, \mu_1, \phi | \sigma, \rho, p, q, Y, H, S) \).

According to the Bayesian rule, the conditional posterior distribution is proportional to the product of the likelihood function and the prior conditional distribution. That means

\[
 f(\mu_0, \mu_1, \phi | \sigma, \rho, p, q, Y, H, S) \propto f(Y | \mu_0, \mu_1, \phi, \sigma, \rho, p, q, H, S) \cdot \\
 f(\mu_0, \mu_1, \phi | \sigma, \rho, p, q, H, S).
\]

In MCMC methods, in order to obtain the posterior distributions that are easily to sample, conjugate priors are often needed. The conjugate prior is a prior that can make the posterior and the prior distributions belong to the same family distribution. We assume a prior normal for \( \beta \equiv (\mu_0, \mu_1, \phi)' \), i.e. \( \beta \sim N_3(B_0, A_0)_{I(\phi < 1 \text{ and } \mu_1 > 0)} \), where \( B_0 \) and \( A_0 \) are the mean and variance of the prior normal and the indicator function \( I(\cdot) \) ensures that the log-volatility process is stationary and the identification of regimes could be guaranteed. This prior is a conjugate prior and the posterior distribution of \( \beta \) is also
normal, i.e. \( \beta|H \sim \mathcal{N}(B_1, A_1)_I(|\phi|<1 \text{ and } \mu_1>0) \), where 

\[
B_1 = (A_0^{-1} + \sigma^{-2}X'X)^{-1}(A_0^{-1}B_0 + \sigma^{-2}X'H^*) \\
A_1 = (A_0^{-1} + \sigma^{-2}X'X)^{-1}, \quad X = (1', (s_2', \ldots, s_{T+1}'), (h_1, \ldots, h_T)') \text{ and } H^* = (h_2, \ldots, h_{T+1})'.
\]

We could sample the random draws from this posterior normal. We specify the hyperparameters values of \(B_0\) and \(A_0\) as \((-2, 1, 0.9)\) and \(20 \cdot I_3\) here, which actually give an uninformative prior. To constrain \(\mu_1 > 0\) and \(|\phi| < 1\), if the generated value of \(\beta\) does not satisfy these conditions, we will discard the draws. Otherwise, we save them.

2. Drawing \(\rho\) and \(\sigma\) from \(f(\rho, \sigma|\mu_0, \mu_1, \phi, p, q, Y, H, S)\).

In this step, we adopt the method proposed by Jacquier et al. (2004), which transform the covariance matrix \(\Sigma\) as

\[
\Sigma = \begin{bmatrix} 1 & \rho \sigma \\ \rho \sigma & \sigma^2 \end{bmatrix} = \begin{bmatrix} 1 & \varphi \\ \varphi & \omega + \varphi^2 \end{bmatrix},
\]

where \(\omega = \sigma^2(1 - \rho^2)\) and \(\varphi = \rho \sigma\). It is easy to get that \(|\Sigma| = \omega\) and

\[
\Sigma^{-1} = \frac{1}{\omega} \begin{bmatrix} \varphi^2 & -\varphi \\ -\varphi & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \frac{1}{\omega} \mathcal{C} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},
\]

where \(\mathcal{C}\) contains \(\varphi\) only. The likelihood function of \((\rho, \sigma^2)\) is readily available as
\[
    l(\rho, \sigma^2) = \prod_{t=1}^{T} f(b_t|\Sigma) \propto |\Sigma|^{-\frac{T}{2}} \exp\left( -\frac{1}{2} \sum_{t=1}^{T} b_t^t\Sigma^{-1}b_t \right)
\]

where \( b_t = (\epsilon_t, \eta_t)' \), \( t = 1, ..., T \). After the transformation of the covariance matrix \( \Sigma \), the likelihood function of \( l(\rho, \sigma^2) \) becomes

\[
    l(\varphi, \omega) \propto \omega^{-\frac{T}{2}} \exp\left( -\frac{1}{2\omega} tr(\Sigma^{-1} \sum_{t=1}^{T} b_t b_t') \right),
\]

where \( R = \sum_{t=1}^{T} b_t b_t' = (e, \eta)'(e, \eta) \) and \( e = (\varepsilon_1, ..., \varepsilon_T)' \) and \( \eta = (\eta_1, ..., \eta_T)' \) are the innovations of the model.

We use conjugate priors such that \( \omega \sim IG(\gamma_0/2, \gamma_1/2) \) and \( \varphi | \omega \sim N(0, \omega/2) \). Then the joint posterior distribution of \((\varphi, \omega)\) can be decomposed into a normal and inverse-gamma distribution. Specifically,

\[
    \varphi | \omega, Y \sim N\left[ \bar{\varphi}, \frac{\omega}{2 + e' e} \right],
\]

where \( \bar{\varphi} = e' \eta / (2 + e' e) \), and
\[
\omega | Y \sim IG\left(\frac{T + 1 + \gamma_0}{2}, \frac{\gamma_1 + \eta'\eta - \frac{(e'\eta)^2}{2 + e'e}}{2}\right).
\]

We define the hyperparameters values of the inverse Gamma as \(\gamma_0 = 2\) and \(\gamma_1 = 0.01\).

In Gibbs sampling, once \(\varphi\) and \(\omega\) are drawn, we can obtain \(\rho\) and \(\sigma\) easily because \(\sigma = \omega + \varphi^2\) and \(\rho = \frac{\varphi}{\sigma}\).

3. Drawing \(h_t (t = 1, \ldots, T)\).

The sampling of the volatilities is a hard part in the Gibbs sampler for the Stochastic Volatility models. In the literature, there are two ways to do it. One is the Single Move approach proposed by Jacquier, Polson and Rossi (1994), the other one is the Multi-move approach proposed by Kim, Shephard and Chib (1998).

(1) Single Move Approach for the Samplings of Volatilities.

In single move samplings, each of the log volatility \(h_t\) is sampled from \(f (h_t | h_{-t}, \theta, Y, S)\) one at a time, where \(h_{-t}\) means all elements of \(h_1, \ldots, h_T\) except \(h_t\).

Because of the Markov property, the conditional distribution \(f (h_t | h_{-t}, Y)\) can be written as \(f (h_t | h_{-t}, Y) \propto f (y_t | h_t) f (h_{t+1} | h_t) f (h_t | h_{t-1})\), where parameters \(Y\) and regime states vector \(S\) are omitted for simplicity. Sampling from \(f (h_t | h_{-t}, Y)\) requires more efforts. First, the constant of proportionality is unknown in the approximation.
\( f(h_t|h_{-t}, Y) \propto f(y_t|h_t)f(h_{t+1}|h_t)f(h_t|h_{t-1}) \). Second, the product of \( f(y_t|h_t)f(h_{t+1}|h_t)f(h_t|h_{t-1}) \) is a non standard distribution which is the product of normal and log normal densities. As mentioned above, the solution is to find a blanketing density which has a standard distribution form and easy to sample. One can sample from \( f(h_t|h_{-t}, Y) \) by drawing samples from the blanketing density and developing a simple accept/rejection or accept/rejection Metropolis-Hastings procedure to decide whether accept the draws from the blanketing density.

The weakness of the single move samplers is they produce a highly correlated sample and are slow to converge, which is showed in Kim, Shephard and Chib (1998). In order to improve the single move samplers, Multi-move sampler is an alternative to it.

(2) Multi-move Approach for the Samplings of Volatilities

Multi-move samplers can sample all latent variables \( H = (h_1, \cdots, h_T) \) at once and can produce less correlated draws and are faster to converge. Cater and Kohn (1994), Shephard (1994), Kim, Shephard and Chib (1998) and Chib, Nardari and Shephard (2002) all suggest transferring the nonlinear and non-Gaussian state space model into a linear and Gaussian state space model where all volatilities can be sampled through using Kalman Filter. In this section, we will adopt this idea and draw the samples for all volatilities in one block in the Gibbs iteration.

After taking the logarithm of the squares of observations, we can have
Observation equation: \[ y_t^* \equiv \log(y_t^2) = h_t + \log(\varepsilon_t^2). \quad (3.1) \]

Treating Eq. (3.1) as the observation equation and Eq. (2.7) as the state equation, we have the form of a state-space model except that \( \log(\varepsilon_t^2) \) is not Gaussian (actually \( \log(\chi_t^2) \)) and the square transformation above fails to retain the correlation between \( \varepsilon_t \) and \( \eta_t \).

To capture the correlated errors, we can write Eq. (2.7) as

\[
h_{t+1} = \mu_0 + \mu_1 S_{t+1} + \phi h_t + \rho \varepsilon_t + \eta_t^*,
\]

where \( \eta_t^* \sim iid \mathcal{N}(0, \sigma^2(1 - \rho^2)) \), and substitute \( \varepsilon_t \) for \( y_t \exp(-h_t/2) \), getting the state equation as

State equation: \[ h_{t+1} = \mu_0 + \mu_1 S_{t+1} + \phi h_t + \rho \sigma y_t \exp(-h_t/2) + \eta_t^*, \mu_1 > 0 \quad (3.2) \]

To solve for the non-Gaussian problem, Kim, Shephard and Chib (1998) uses a mixture of seven normal distributions to approximate the distribution of \( \log(\varepsilon_t^2) \), while Ompri, Chib, Shephard and Nakajima (2007) improves the approximation by using the mixture of 10 normal distributions. Here, we choose the 10-mixture approximation, i.e.

\[
f(\log(\varepsilon_t^2)) \approx \sum_{i=1}^{K=10} p_i \mathcal{N}(m_i, \sigma_i^2),
\]
where $p_i, m_i$ and $\sigma_i^2$ are given in table 3.1.\footnote{The method of Kim, Shephard and Chib (1998) is to select K and the values of $p_i, m_i$ and $\sigma_i^2$ in order to match the first four moments of the mixture approximation and the $\log (\chi^2_i)$ distribution. Ompri, Chib, Shephard and Nakajima (2007) favor a tighter approximation to the $\log (\chi^2_i)$ distribution increasing K from 7 to 10. They all check the difference between the $\log (\chi^2_i)$ distribution and the approximating mixture distribution to ensure that the approximation is “sufficiently good.”}

Table 3.1 Selection of the mixing distribution for approximating $\log(\epsilon_i^2)$

<table>
<thead>
<tr>
<th>Component $i$</th>
<th>Probability $p_i$</th>
<th>Mean $m_i$</th>
<th>Variance $\sigma_i^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00609</td>
<td>1.92677</td>
<td>0.11265</td>
</tr>
<tr>
<td>2</td>
<td>0.04775</td>
<td>1.34744</td>
<td>0.17788</td>
</tr>
<tr>
<td>3</td>
<td>0.13057</td>
<td>0.73504</td>
<td>0.26768</td>
</tr>
<tr>
<td>4</td>
<td>0.20674</td>
<td>0.02266</td>
<td>0.40611</td>
</tr>
<tr>
<td>5</td>
<td>0.22715</td>
<td>-0.85173</td>
<td>0.62699</td>
</tr>
<tr>
<td>6</td>
<td>0.18842</td>
<td>-1.97278</td>
<td>0.98583</td>
</tr>
<tr>
<td>7</td>
<td>0.12047</td>
<td>-3.46788</td>
<td>1.57469</td>
</tr>
<tr>
<td>8</td>
<td>0.05591</td>
<td>-5.55246</td>
<td>2.54498</td>
</tr>
<tr>
<td>9</td>
<td>0.01575</td>
<td>-8.68384</td>
<td>4.16591</td>
</tr>
<tr>
<td>10</td>
<td>0.00115</td>
<td>-14.65000</td>
<td>7.33342</td>
</tr>
</tbody>
</table>

Ompri, Chib, Shephard and Nakajima (2007) show the evidence that the $K = 10$ components lead to a superior approximation to $K = 7$ components.
The observation equation (3.1) now could be regarded as

\[ y_t^* = c_t + h_t + v_t, \quad v_t \sim iid \mathcal{N}(0, Q_t), \]  

(3.3)

where \((c_t, Q_t)\) assumes the value \((m_i, \sigma_i^2)\) of Table 3.1 for some \(i\). Then Eq. (3.3) and Eq. (3.2) construct a Gaussian but non-linear state space model. For this special state space model, we will use the extended Kalman filter and backward sampling method (Fruhwirth-Schnatter(1994) and Carter and Kohn (1994)) to generate the series \((h_1, \cdots, h_T)\) jointly at a time (i.e. Multi-move sampler).

Firstly, we need to determine \(c_t\) and \(Q_t\). Augment the model with a series of independent indicator variables \(\{I_t\}\), where \(I_t\) assumes a value in \(\{1, 2, \cdots, 10\}\) such that \(P(I_t = i) = p_{it}\) with \(\sum_{t=1}^{10} p_{it} = 1\) for each \(t\). Probability \(p_i\) in Table 3.1 is the prior distribution of \(I_t\), while the posterior distribution of \(I_t\) is

\[ p_{it} = \frac{p_i q_{it}}{\sum_{j=1}^{10} p_j q_{jt}}, \quad i = 1, \ldots, 10, \]

where \(q_{it}\) are the likelihood functions of \(I_t\) and \(q_{it} = \phi[(y_t^* - h_t - m_i)/\sigma_i]\), for \(i = 1, \ldots, 10\), with \(\phi(.)\) being the p.d.f. of the standard normal distribution. The sampling of \(I_t\) is realized by the inverse transform method and the values of \(c_t\) and \(Q_t\) are then determined.
Here, the extended Kalman filter (EKF), an approximate nonlinear Bayesian filter, for this nonlinear state-space model is specified as:

\[
\begin{align*}
    v_t &= y^*_t - y^*_{t|t-1} = y^*_t - c_t - h_{t|t-1}, \\
    V_t &= \Sigma_{t|t-1} + Q_t, \\
    h_{t|t} &= h_{t|t-1} + \Sigma_{t|t-1} V_t^{-1} v_t, \\
    \Sigma_{t|t} &= \Sigma_{t|t-1} - \Sigma_{t|t-1} V_t^{-1} \Sigma_{t|t-1}, \\
    h_{t+1|t} &= G(h_t), \\
    \Sigma_{t+1|t} &= g(h_{t|t})^2 \Sigma_{t|t} + \sigma^2 (1 - \rho^2),
\end{align*}
\]  

(3.4)

where \( V_t = Var(v_t) \) is the variance of the 1-step ahead prediction error, i.e. \( v_t \), of \( y^*_t \) given \( F_{t-1} = (y^*_1, ..., y^*_t) \), and \( h_{j|i} \) and \( \Sigma_{j|i} \) are, respectively, the conditional expectation and variance of the state variable \( h_j \) given \( F_i \). The specific function form of \( G(\cdot) \) is from Eq. (3.2), i.e. \( G(h_t) = \mu_0 + \mu_1 S_{t+1} + \phi h_t + \rho \sigma y_t \exp(-h_t/2) \) and \( g(h_t) \) is the first-order derivative of \( G(h_t) \).

Let \( f(h_1, \cdots, h_T | F_T) \) represent the joint posterior distribution of log-volatilities given the whole dataset and other parameters, where for simplicity other parameters are omitted. We could partition this joint density as
\[ f(h_1, \ldots, h_T|F_T) \]
\[ = f(h_T|F_T)f(h_{T-1}|h_T, F_T)f(h_{T-2}|h_{T-1}, h_T, F_T) \cdots f(h_1|h_2, h_3, \ldots, h_T, F_T) \]
\[ = f(h_T|F_T)f(h_{T-1}|h_T, F_T)f(h_{T-2}|h_{T-1}, F_T) \cdots f(h_1|h_2, F_T) \]
\[ = f(h_T|F_T)f(h_{T-1}|h_T, F_{T-1})f(h_{T-2}|h_{T-1}, F_{T-2}) \cdots f(h_1|h_2, F_1) \]

The second equality holds because the log-volatility process is a Markov process, i.e. conditional on \( h_{t+1}, h_{t+j} \) is independent of \( h_t \) for \( j>1 \).

The whole vector \((h_1, \cdots, h_T)'\) can be jointly generated by a backward method using quantities available from the Extended Kalman Filter (EKF), i.e. Eq.\((3.4)\). Specifically, we firstly generate \( h_T \) from \( h_T|F_T \sim \mathcal{N}(h_{T|T}, \Sigma_{T|T}) \). And then for \( t = T - 1, T - 2, \cdots, 1 \), we can generate \( h_t \) from \( f(h_t|h_{t+1}, F_t) \).

From the Extended Kalman Filter (EKF), \( f(h_t, h_{t+1}|F_t) \) is a bivariate normal as

\[
\begin{pmatrix}
  h_t \\
  h_{t+1}
\end{pmatrix}_{F_t} \sim \mathcal{N}\left(\begin{pmatrix}
  h_{t|t} \\
  h_{t+1|t}
\end{pmatrix}, \begin{pmatrix}
  \Sigma_{t|t} & \phi \Sigma_{t|t} \\
  \phi \Sigma_{t|t} & \Sigma_{t+1|t}
\end{pmatrix}\right).
\]

By Theorem for multi-normal distributions, \( f(h_t|h_{t+1}, F_t) \sim \mathcal{N}(\mu^*_t, \Sigma^*_t) \), where

\[
\mu^*_t = h_{t|t} + \phi \Sigma_{t|t} \Sigma_{t+1|t}^{-1}(h_{t+1} - h_{t+1|t}).
\]
\[
\Sigma^*_t = \Sigma_{t|t} - \phi^2 \Sigma_{t|t} \Sigma_{t+1|t}^{-1}.
\]
4. Drawing $S_t (t = 1, \cdots, T)$.

In this section, we also use the multi-move sampler to draw the sequence of $\{S_t\}$, i.e. generating the samples from the join distribution $g (S_1, \cdots, S_T | \theta, H)$ at once. In this step, the Hamilton filter by Hamilton in 1989 and the backward sampling method by Fruhwirth-Schnatter in 1994 and Carter and Kohn in 1994 will be used.

According to the Markov property in the process of regime states variables, the above joint distribution could be composed to

$$g(S_1, \cdots, S_T | \theta, H) = g(S_T | \theta, H_T) \prod_{t=1}^{T-1} g(S_t | S_{t+1}, \theta, H_t),$$

where $H_t = (h_1, \cdots, h_t)$. As suggested by the above expression, we can generate first $S_T$ from $g(S_T | \theta, H_T)$, and then for $t = T-1, T-2, \cdots, 1$, we generate $S_t$ from $g(S_t | S_{t+1}, \theta, H_t)$ recursively. For this purpose, we firstly run Hamilton filter and get the values $g(S_t | \theta, H_t)$ for each time $t$.

The specific Hamilton filter is defined as

$$g(S_t = j | \theta, H_t) = \frac{g(S_t = j | \theta, H_{t-1}) \Pr(H_t = j | S_t = j, \theta, H_{t-1})}{\sum_{j=0}^{1} g(S_t = j | \theta, H_{t-1}) \Pr(H_t = j | S_t = j, \theta, H_{t-1})}.$$
\[ g(S_{t+1} = j | \theta, \bar{H}_t) = \sum_{i=1}^{0} P_{ij} g(S_t = i | \theta, \bar{H}_t), \quad i, j = 0 \text{ or } 1, \]

where \( P_{ij} \) represents the probability of transitions from \( S_t = i \) to \( S_{t+1} = j \).

The last iteration of the Hamilton filter provides \( g(S_T | \theta, \bar{H}_T) \) from which \( S_T \) can be generated. To generate \( S_t \) for \( t = T - 1, T - 2, \ldots, 1 \), we employ the following results,

\[
\frac{g(S_t | S_{t+1}, \theta, \bar{H}_t)}{g(S_t, S_{t+1} | \theta, \bar{H}_t)} = \frac{g(S_{t+1} | S_t, \theta, \bar{H}_t) g(S_t | \theta, \bar{H}_t)}{g(S_{t+1} | \theta, \bar{H}_t)}.
\]

where \( g(S_{t+1} | S_t, \theta, \bar{H}_t) \) is the transition probability, and \( g(S_t | \theta, \bar{H}_t) \) is readily available from the above Hamilton filter. We select the value of 0 or 1 for the state variables based on draws from uniform distribution.

5. Drawing \( p \) and \( q \).

Beta distributions are used here as the conjugate priors for the transition probabilities. The priors for \( p \) and \( q \) are \( q \sim beta(u_{00}, u_{01}) \) and \( p \sim beta(u_{11}, u_{10}) \) respectively, where we set the hyperparameters of the priors equal to different values and observe that the results are quite similar across these experiments. The conditional likelihood is \( l(p, q) = q^{n_{00}} (1 - q)^{n_{01}} p^{n_{11}} (1 - p)^{n_{10}} \), where \( n_{ij} \) is equal to the number of transition from state \( i \) to state \( j \), \( i \) or \( j = 0, 1 \). The resulting posteriors are also beta distributions like
\( q \sim \text{beta}(u_{00} + n_{00}, u_{01} + n_{01}) \) and \( p \sim \text{beta}(u_{11} + n_{11}, u_{10} + n_{10}) \) and then the random samples from these distributions could be drawn.

We initialize \( \theta, S \) and \( H \), and then run the above five steps, the one cycle or swept for the Gibbs sampler. We can do the same iteration many times and regard the samples from this sampler as the approximately random samples from the joint posterior distributions \( f(\theta, H, S \mid Y) \) when the iteration times are large enough. These samples then could be used to estimate the marginal density of each parameter, obtain the inference to the parameters, like mean and variance.
4. Monte Carlo Simulation Studies for SVLR Model

The simulation study is a numerical technique for conducting experiments on the computer. The Monte Carlo simulation is a computer experiment involving random sampling from probability distributions. Usually, when statisticians talk about “simulations”, they mean “Monte Carlo simulations”.

Properties of our estimation methods need to be established so that the methods could be used with confidence. In addition, the consequences of using a mis-specified model, such as using the basic stochastic volatility model when the leverage effect or regime switching are present, need to be studied. Monte Carlo Simulation studies could help us to fulfill these purposes. In this chapter, Monte Carlo sampling experiments will be conducted to gauge the performance of Bayesian parameter estimation and volatility estimation.

4.1 Performance of Bayesian Parameter Estimation

Knowledge of the sampling distribution is the key to evaluate the behavior of a statistic. For example, a researcher can determine the bias of a statistic from the sampling distribution, as well as its efficiency and other desirable properties. Sampling distributions are theoretical and unobserved, however, the Monte Carlo simulations could be used to approximate the sampling distributions of estimators or statistics.
In our Monte Carlo studies, data are generated from the population described by our SVLR model with hypothesized parameter values. We set $\phi = 0.7, \mu_0 = -4, \mu_1 = 5, \sigma = 1.9, \rho = -0.3, \ q = 0.75$ and $p = 0.95$. We specify the length of the time series as $T = 1000$. Using the true transition possibilities $p$ and $q$, we generate a state vector $(S_1, \cdots, S_T)$ with elements 0 or 1. Using the state vector and the true values for parameter $\theta$, we can generate the log volatilities $(h_1, \cdots, h_T)$. Then we can generate the simulated dataset $Y = (y_1, \cdots, y_T)$.

A large number of datasets could be generated through the way mentioned above. The number of samples to be drawn (replications) can be thought of as the sample size for the Monte Carlo study. The number of replications should be increased until stability of the results is achieved. Another technical consideration in the Monte Carlo study is the seed. The value of the seed determines the starting point for the random draws of the samples.

In our Monte Carlo simulation studies to the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model), We simulate 500 samples of datasets and $T = 1000$ for each dataset. For each simulated dataset, we generate 20,000 draws from the MCMC Gibbs sampling algorithm, discard the first 10,000 draws and record the last 10,000 ones, i.e. the number of burn-in iterations is 10,000. The posterior parameter estimators are computed by averaging the last retained 10,000 draws. Finally, we could get 500 posterior estimates for each parameter.
The performance of parameter estimations could be judged from the results contained in Table 4.1. The row labeled “Average” reports the average of the posterior parameter estimates over the 500 replications in the Monte Carlo study. The row labeled “RMSE” reports the Root Mean Squared Errors (RMSE) of the 500 posterior means. The Root Mean Squared Errors are defined as \( \sqrt{\frac{1}{K} \sum_{k=1}^{K} (\hat{\theta}_k - \theta_0)^2} \), where \( K \) is the number of samples simulated in the Monte Carlo studies, equal to 500 in our studies, \( \hat{\theta}_k \) is the posterior estimate obtained from the \( k \)th simulated data and \( \theta_0 \) is the hypothesized parameter values for the population.

All of the parameters except \( \rho \) are recovered precisely. The estimator \( \rho \) appears to have bias. Actually Bayes estimators do not have to be unbiased. However, as the data length increases this bias should decrease. Indeed, we experimented with data length \( T=5000 \) and found that the bias in estimation of \( \rho \) can decrease. The purpose of sample size \( T=1000 \) here is to save time and represent a “stress-test” of our method.

### Table 4.1 Sampling properties of Bayes estimators, SVLR model

<table>
<thead>
<tr>
<th></th>
<th>( \phi )</th>
<th>( \sigma )</th>
<th>( \rho )</th>
<th>( \mu_0 )</th>
<th>( \mu_1 )</th>
<th>( q )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True values</strong></td>
<td>0.7</td>
<td>1.8</td>
<td>-0.3</td>
<td>-4</td>
<td>5</td>
<td>0.95</td>
<td>0.75</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>0.68</td>
<td>1.99</td>
<td>-0.15</td>
<td>-4.14</td>
<td>4.98</td>
<td>0.96</td>
<td>0.77</td>
</tr>
<tr>
<td><strong>RMSE</strong></td>
<td>0.03</td>
<td>0.22</td>
<td>0.15</td>
<td>0.35</td>
<td>0.27</td>
<td>0.01</td>
<td>0.05</td>
</tr>
</tbody>
</table>
The evidence in Table 4.1 shows that the Bayes estimators obtained from our MCMC Gibbs sampler have good sampling properties for the SVLR model. The sampling performance provides indirect evidence that our Gibbs sampler algorithm could work well with a moderate number of draws. More details about the convergence of our Gibbs sampler will be explained in the future section.

We can check the performance of parameter estimators under mis-specification. We generate the simulated data from the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) but fit the basic Stochastic Volatility Model, the Stochastic Volatility Model with Leverage Effect (SVL model) respectively. In this simulation studies, we also simulate 500 sampled datasets with 1000 observations for each. The posterior estimators are also gained from the last retained 10,000 draws after discarding the first 10,000 burn-in draws in the MCMC Gibbs sampler. The Gibbs sampler algorithms for the basic SV model and SVL model are modified from our algorithm for the SVLR model. Table 4.2 presents the effects of estimating the wrong models on parameter estimations. We could see that if the data are generated from the process described by SVLR model, use of the basic SV model or the SVL model will lead to substantial biases to the parameter estimation. If the fit model could not incorporate the structure change in the volatility process, the estimates of persistence also the autocorrelation coefficient $\phi$ will be spuriously high. Actually part of the persistence should come from the fact that the system has the potential to remain in the same regime. This finding implies that examining the existence of switching regimes for a volatility process is important.
Table 4.2 Estimating wrong models: effect on parameter estimates

<table>
<thead>
<tr>
<th>Data generated from the SVLR model</th>
<th>( \phi )</th>
<th>( \sigma )</th>
<th>( \rho )</th>
<th>( \mu_0 )</th>
<th>( \mu_1 )</th>
<th>( q )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Value</td>
<td>0.7</td>
<td>1.8</td>
<td>-0.3</td>
<td>-4</td>
<td>5</td>
<td>0.95</td>
<td>0.75</td>
</tr>
<tr>
<td><strong>Basic SV model fit</strong></td>
<td>( \phi )</td>
<td>( \sigma )</td>
<td>( \mu )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average RMSE</td>
<td>0.86</td>
<td>2.63</td>
<td>-1.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SVL model fit</strong></td>
<td>( \phi )</td>
<td>( \sigma )</td>
<td>( \rho )</td>
<td>( \mu )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average RMSE</td>
<td>0.16</td>
<td>0.83</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.2 Performance of Bayesian Volatility Estimation

Volatility estimation and prediction is one of the most uses of the SV models. In this section, we will study the performance of SVLR model for estimating volatilities and the consequence of using wrong models on the estimates of volatility. For each observation of a simulated dataset, we can examine the performance of Bayes estimates of \( \exp(h_t/2) \). The Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) are calculated through averaging over the 1000 observations of the 500 simulated datasets. The Mean Absolute Error is defined as the average of the absolute error, \( |\exp(h_t/2) - \exp(h_t/2)| \), over all observations. Table 4.3 compares the performance of volatility
estimation of the SVLR model, the basic SV model and the SVL model. Smaller values for RMSE and MAE mean better performance. We could see that for the data generated from the SVLR model, the wrong use of basic SV model or SVL model would impair the volatility estimation.

Table 4.3 Estimating wrong models: effect on the estimated smoothing volatilities

<table>
<thead>
<tr>
<th></th>
<th>Data are generated from the SVLR model</th>
<th>All obs. (1000 x 500)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SVLR model fit</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE (exp(h_t/2))</td>
<td></td>
<td>1.8854</td>
</tr>
<tr>
<td>MAE (exp(h_t/2))</td>
<td></td>
<td>0.1966</td>
</tr>
<tr>
<td><strong>Basic SV model fit</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE (exp(h_t/2))</td>
<td></td>
<td>1.9132</td>
</tr>
<tr>
<td>%MAE (exp(h_t/2))</td>
<td></td>
<td>0.2002</td>
</tr>
<tr>
<td><strong>SVL model fit</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE (exp(h_t/2))</td>
<td></td>
<td>2.0356</td>
</tr>
<tr>
<td>%MAE (exp(h_t/2))</td>
<td></td>
<td>0.2036</td>
</tr>
</tbody>
</table>
5. The Empirical Application of SVLR Model to Stock Returns

After doing the Monte Carlo simulation studies in chapter 4, we could be confident of the properties of our MCMC Gibbs sampler methods for the studies to the SVLR model. In this chapter, we will apply the SVLR model to the real stock return data, estimate the parameters as well as the latent log volatility and regime state variables.

5.1 Data Description

To illustrate the performance of the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) to fit the actual stock return data and compare findings from the SVLR model with findings in the existing SV models, we will analyze the S&P 500 index, which is widely used as a measurement of the general level of stock prices. The series we use for illustration is the S&P 500 index weekly returns from the first week of April in 1986 to the last week of December in 2011. Wednesday-to-Wednesday returns are computed. If the Wednesday index is not available, the Thursday or Friday index will be substituted. There are 1342 observations totally. The use of these weekly returns could minimize the effects of weekend, holidays, and day-of-the-week on the analysis while still remains the information present in the daily data. Table 5.1 provides summary statistics for the series \( y_t, y_t^2 \) and \( \log(y_t^2) \), where \( y_t \) is the S&P 500 weekly return at time \( t \).
Table 5.1 Summary statistics for the S&P 500 index weekly return (%)

<table>
<thead>
<tr>
<th></th>
<th>$y_t$</th>
<th>$y_t^2$</th>
<th>$\log(y_t^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.0013</td>
<td>0.00058</td>
<td>-9.059</td>
</tr>
<tr>
<td>S.D.</td>
<td>0.024</td>
<td>0.0017</td>
<td>2.286</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.832</td>
<td>13.09</td>
<td>-1.122</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>9.265</td>
<td>270.129</td>
<td>5.069</td>
</tr>
</tbody>
</table>

**Autocorrelations**

<table>
<thead>
<tr>
<th>Lags</th>
<th>$y_t$</th>
<th>$y_t^2$</th>
<th>$\log(y_t^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.0498</td>
<td>0.2984</td>
<td>0.1035</td>
</tr>
<tr>
<td>2</td>
<td>0.0544</td>
<td>0.1389</td>
<td>0.0902</td>
</tr>
<tr>
<td>3</td>
<td>-0.0583</td>
<td>0.1797</td>
<td>0.1583</td>
</tr>
<tr>
<td>4</td>
<td>-0.0313</td>
<td>0.0999</td>
<td>0.0931</td>
</tr>
</tbody>
</table>

Note: Kurtosis refers to excess kurtosis. The period time for the data is from 04/1986 to 12/2011.

All of the series, $y_t$, $y_t^2$ and $\log(y_t^2)$, have excess kurtosis greater than 0. The autocorrelations of $y_t$ are small, lying in the 95% confidence interval [-0.059,0.059] and showing that they are significantly close to zero. The autocorrelations of $y_t^2$ and $\log(y_t^2)$ are much higher than those of $y_t$. This shows that the transformed series, $y_t^2$ and $\log(y_t^2)$, are significantly autocorrelated and that the AR(1) specification for the log volatility process makes sense.
5.2 Estimation Results of SVLR Model to Actual Data

We perform our MCMC Gibbs sampler algorithm on the S&P 500 index weekly return. We generate 100,000 samples from the Gibbs sampler, the first 50,000 burn-in iterations are discarded and the retained 50,000 iterations are regarded as the simulated samples from the posterior density. Here the large iteration times and burn-in size are enough to remove the effects of initial starting values on the performance of the Gibbs sampler.

5.2.1 Convergence of Gibbs Sampler

In order to detect convergence of the Gibbs sampling algorithm, the method similar to that of McCulloch and Rossi (1994) is adopted, where the empirical distributions of the simulated draws are compared when the Gibbs sampler is initiated from different starting points, and as the number of sampling draws increases, the changes in these distributions are trivial.

Another standard test of convergence of Gibbs samplers, the autocorrelation function of the drawn sequences, is shown in Figure 5.1. We could see that the autocorrelations decay quickly for most of the parameters, which imply that the simulated draws from our Gibbs sampler could converge relatively quick. In addition, the bell-like shapes of histograms of simulated draws from the sampler in Figure 5.2 can also imply the convergence of the Markov chain sampler.
Figure 5.1 Estimation results for S&P 500 data –
Sample Autocorrelation Functions of simulated draws from the Gibbs sampler.
5.2.2 Estimation Results

The posterior means, Monte Carlo standard error, posterior 95% credible interval and inefficiency factors are displayed in Table 5.2. The drawn samples from the Gibbs sampler can be used to make the inferences of the SVLR model such as estimating the posterior moments or marginal densities of parameters. The posterior means can be calculated by averaging the simulated draws from the Gibbs sampler and be regarded as the estimates of the parameters.
It should be kept in mind that the drawn samples from the MCMC algorithm are correlated. This serial correlation can be very high for badly designed sampler. The standard error (S.E.) here means the Monte Carlo Standard Error (MC S.E.), a numerical standard error representing the accuracy of the resulting estimates. It can be estimated by time series methods to account for the serial correlation in the simulated draws. Specifically, we adopt the covariance method to estimate the Monte Carlo Standard Error.\textsuperscript{8}

The 2.5th and 97.5th percentiles of the simulated draws specify the posterior 95% credible intervals. A credible interval (or Bayesian confidence interval) in Bayesian statistics is an interval in the domain of a posterior probability distribution. If the probability that the parameter lies in a range is 0.95, this range is called the 95% credible intervals. Credible intervals are analogous to confidence intervals in frequentist statistics, but they have different meanings. A frequentist 95% confidence interval means that with a large number of repeated samples, 95% of the calculated confidence intervals would include the true value of the parameter. In frequentist terms, the parameter is fixed (cannot be considered to have a distribution of possible values) and the confidence interval is random (it depends on the random sample). In general, Bayesian credible intervals do not coincide with frequentist confidence intervals since credible intervals incorporate problem-specific contextual information from the prior distribution whereas confidence intervals are based only on the data.

\textsuperscript{8} Suppose that $X_1, X_2, \ldots, X_N$ are simulated samples from the Gibbs sampler for parameter $X$, the posterior mean of the parameter is the sample average $\bar{X} = N^{-1} \sum_{i=1}^{N} X_i$. For large $N$, the variance of $\bar{X}$ can be estimated by $\frac{S^2}{N}$, where $S^2 = \hat{R}(0) + 2 \sum_{t=1}^{K} \hat{R}(t)$ and $\hat{R}(0), \ldots, \hat{R}(K)$ are sampled (auto)covariances at lag $K$. The Monte Carlo standard error would the square root of the estimate of $\text{Var}(\bar{X})$, i.e., $\sqrt{\text{Var}(\bar{X})}$.\vspace{1cm}
The Inefficiency Factor serves to quantify the relative efficiency loss in the computation of the posterior means from correlated versus independent samples. Here the inefficiency factor is \( 1 + 2 \sum_{s=1}^{500} \hat{\rho}_s \), where \( \hat{\rho}_s \) is the sample autocorrelation at lag \( s \) calculated from the generated draws of our Gibbs sampler and we let the maximum lag be 500. The relatively low values of inefficiency factors in our results show that our algorithm for estimation is efficient enough.

Table 5.2 Estimation Results of SVLR Model for S&P 500 Data

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>MC S.E.</th>
<th>95% Credible Interval</th>
<th>Inefficiency Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi )</td>
<td>0.67</td>
<td>0.0036</td>
<td>[0.467, 0.798]</td>
<td>91.34</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.53</td>
<td>0.0021</td>
<td>[0.432, 0.647]</td>
<td>73.10</td>
</tr>
<tr>
<td>( \rho )</td>
<td>-0.30</td>
<td>0.0011</td>
<td>[-0.400, -0.181]</td>
<td>18.49</td>
</tr>
<tr>
<td>( \mu_0 )</td>
<td>0.18</td>
<td>0.0029</td>
<td>[0.069, 0.327]</td>
<td>98.99</td>
</tr>
<tr>
<td>( \mu_1 )</td>
<td>0.43</td>
<td>0.0056</td>
<td>[0.230, 0.768]</td>
<td>84.81</td>
</tr>
<tr>
<td>( q )</td>
<td>0.993</td>
<td>0.0001</td>
<td>[0.983, 0.999]</td>
<td>15.85</td>
</tr>
<tr>
<td>( p )</td>
<td>0.994</td>
<td>0.0001</td>
<td>[0.982, 0.999]</td>
<td>51.09</td>
</tr>
</tbody>
</table>
The persistence parameter $\phi$ for the AR (1) log-volatility process has posterior mean equal to 0.67. We could compare this AR (1) coefficient estimate fitted from the Stochastic Volatility Model with Leverage Effect and Regime Switching Model (SVLR model) with the one from the Stochastic Volatility Model with only leverage effect but no regime switching (SVL model). Table 5.3 reports the estimation results of SVL model for the same S&P 500 index returns. These results are consistent the previous studies by So et al. (1998), Kalimipalli and Susmel (2004) and Shibata and Watanabe (2005) which document that the estimate of $\phi$ is smaller in the Markov switching SV model than those in the SV model with no regime switching. Usually the high persistence of log volatility process is overestimated. Part of the high persistence of volatility should be explained by the persistence of the volatility to remain in a regime. The estimates of transition possibilities $p$ and $q$ are both very close to one, indicating that the probability of log volatility retaining in the same regime is very high. This result captures an important feature of financial data, the volatility clusters, i.e. volatility may be high for certain periods and low for other periods.

**Table 5.3 Estimation results of SVL model for S&P 500 data**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\phi$</th>
<th>$\sigma$</th>
<th>$\rho$</th>
<th>$\mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.88</td>
<td>0.44</td>
<td>-0.30</td>
<td>0.15</td>
</tr>
<tr>
<td>MC S.E.</td>
<td>0.0011</td>
<td>0.0024</td>
<td>0.0014</td>
<td>0.0015</td>
</tr>
</tbody>
</table>
The posterior mean of $\rho$ is -0.29 and negative as expected, suggesting the presence of leverage effect. Since its 95% credible interval is [-0.40, -0.18], the posterior probability that $\rho$ is negative is greater than 0.95. The difference between the intercept levels of log volatility processes for different states is $\mu_1 = 0.43$, whose 95% credible intervals are [0.23, 0.77], away from 0, showing the presence of different regime states. Since $h_t = \log(\sigma_t^2)$, then the conditional variance $\sigma_t^2$ has a Log-normal distribution. At low-volatility regime the mean level of log-volatility $h_t$ is 0.545 (i.e. $\mu_0/(1 - \phi)$), while at high-volatility regime the mean level of log-volatility $h_t$ is 1.848 (i.e. $(\mu_0 + \mu_1)/(1 - \phi)$). The variance of $h_t$ is equal to 0.51 (i.e. $\sigma^2/(1 - \phi^2)$). The mean level of $\sigma_t^2$ is equal to $\exp(\mu_h + \sigma_h^2/2)$. So, the mean level of $\sigma_t^2$ at high-volatility regime is equal to 8.19, whereas the mean level of $\sigma_t^2$ at low-volatility regime is equal to 2.23.

Smoothed estimates of log-volatility $h_t(t = 1, \cdots, T)$ can be calculated by averaging the last 10,000 draws of the log volatility vector $H$ from the Gibbs sampler. Similarly the smoothed probabilities of high-volatility state can be calculated by averaging the last 10,000 draws of the state vector $S$. Figure 5.3 provides the plots of estimated smoothed volatilities $\exp(h_t/2)$, posterior probabilities of high-volatility state and the S&P 500 weekly returns. We may define period $t$ as a turning point if the posterior probability $P(S_{t-1} = 1|Y) > 0.5$ and $P(S_t = 1|Y) < 0.5$ or $P(S_{t-1} = 1|Y) < 0.5$ and $P(S_t = 1|Y) > 0.5$. Then, the high-volatility periods are around 04/1986 – 08/1988, 09/1989 – 06/1990, 01/1997 – 04/2003 and 06/2007 – 12/2011.
Figure 5.3 Estimation results for S&P 500 data

(a) Estimated smoothed volatility, \( \exp(h_t/2) \).

(b) Posterior probability of being at high-volatility state \( (S_t = 1) \).
5.3 Estimation Results From the Reweighting MCMC Algorithm Corrected For Approximation

In our Gibbs sampling algorithm introduced in chapter 3, we approximate the density of $\log \chi^2$ with the 10-normals mixture density, which provides a good connection for our model and Gaussian state space models and leads to an efficient sampling procedure, as shown above. Thus, the draws from our MCMC procedure, i.e. $(\theta^g, H^g, S^g), g = 1, \cdots, G$, are from the approximate posterior density of parameters and latent variables denoted by $k(\theta, H, S|Y^*)$. In this section, we could show that it is possible to produce draws from the exact correct posterior density of parameters and latent variables, i.e.
\[ f(\theta, H, S|Y), \] through a reweighting step for our MCMC procedure. Kim et al. (1998) propose the general principle for the reweighting schedule.

Define

\[ w(\theta, H, S) = \log f(\theta, H, S|Y) - \log k(\theta, H, S|Y^*) \]

\[ = \text{const} + \log f(Y|H) - \log k(Y^*|H), \]

where

\[ f(Y|H) = \prod_{t=1}^{T} f_N(y_t|0, \exp(h_t)), \]

and

\[ k(Y^*|H) = \prod_{t=1}^{T} \sum_{i=1}^{10} q f_N(y^*_t|h_t + m_i, \omega^2_t). \]

Both of these functions involve Gaussian densities and straightforward to evaluate for any value of \( h \). Then

\[ Eg(\theta)|Y = \int g(\theta)f(\theta|Y)d\theta = \frac{\int g(\theta) \exp(w(\theta, H, S)) k(\theta, H, S|Y^*) d\theta dH dS}{\int \exp(w(\theta, H, S)) k(\theta, H, S|Y^*) d\theta dH dS}, \]

where the denominator is actually 1. Thus, we can now get a sample from the exact posterior density by resampling the samples from the approximate posterior density with weights \( c^\theta \), and
Furthermore, the posterior mean now could be computed by weighted averaging our MCMC draws in the above procedure.

If the mixture approximation is very good, we would expect that the weights $c^g$ have a very small variance. In order to justify the quality of our approximation, we next see the dispersion of the weights. We record the weights from the sampler in section 5.2 and plot the distribution of $\log (c^g \times G)$ in Figure 5.4. The log-weights are close to being normally distributed with a mean of being -0.01 and a standard deviation of being 0.15. A corresponding approximating normal density with the fitted mean and standard deviation are also given in the figure.

The distribution of log-weights demonstrates that reweighting would have a very modest effect on the posterior inference for the parameters. So we would expect that the improvement for the parameter estimates from the reweighting schedule would be very small. In Table 5.4, we report the weighted average of the samples we got from the MCMC procedure in section 5.2. From the table, we could see that the effect of reweighting on the parameter estimations is very trivial and the approximation used in our MCMC procedure is very good.
Table 5.4 Estimation results from the reweighing schedule for the S&P 500 data

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Reweighted</th>
<th>Un-weighted</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi$</td>
<td>0.6688</td>
<td>0.6704</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.5321</td>
<td>0.5294</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.2950</td>
<td>-0.2941</td>
</tr>
<tr>
<td>$\mu_0$</td>
<td>0.1779</td>
<td>0.1784</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>0.4296</td>
<td>0.4292</td>
</tr>
<tr>
<td>$q$</td>
<td>0.9935</td>
<td>0.9935</td>
</tr>
<tr>
<td>$p$</td>
<td>0.9943</td>
<td>0.9942</td>
</tr>
</tbody>
</table>

Figure 5.4 Histogram of the $\log(c^g \times G)$
6. Filtering, Diagnostics and Model Comparison

In this chapter, we will complete the methodological development for the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model). Firstly, we will develop a particle filter to perform filtering, that is computing features of the filtering density \( f(S_t, h_t|Y_t, \theta) \), where \( Y_t = (y_1, \ldots, y_t) \). Knowledge of filtering density will enable us to calculate the diagnostics statistics, marginal likelihoods and Bayes factors. We could do model comparison between our model and the existing models using these results.

6.1 Filtering

The most important problem in the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) is to estimate latent variables, i.e. \( h_t \) and \( S_t \), given the data and the model. Smoothing is to estimate them given the whole dataset \( Y_T = (y_1, \ldots, y_T) \). Actually we have obtained the smoothing estimates for the latent variables from the output of the MCMC algorithm designed in Chapter 3. Filtering is to estimate latent variables given the data throughout the \( t \), i.e. \( Y_t = (y_1, \ldots, y_t) \).

6.1.1 Particle Filter for Filtering in SVLR model

In order to infer properties of the filtering density \( f(S_t, h_t|Y_t, \theta) \), we develop an algorithm called particle filter, which is to sample from the filtering density sequentially from \( t = 0 \) (see Pitt and Shephard (1999)). Our particle filter is the extensions of the
particle filter for the basic SV model proposed by Kim et al. (1998) and the particle filter for the dynamic Markov switching factor model proposed by Kaufmann (2000) and Watanabe (2003). Throughout $\theta$ will be assumed known. In the following studies, we will set $\theta$ to the Bayesian posterior estimates.

Suppose that we have $M$ draws $(h_{t-1}^{(m)}, S_{t-1}^{(m)}) (m = 1, \cdots, M)$ called particles sampled from $f(S_{t-1}, h_{t-1} | Y_{t-1})$, then the filtering density of the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) can be written as

$$f(S_t, h_t | Y_t) \propto f(y_t | h_t) \cdot \int \left( \int f(h_t | S_t, h_{t-1}) \cdot P(S_t | S_{t-1}) \cdot f(S_{t-1}, h_{t-1} | Y_{t-1}) \, dh_{t-1} \, dS_{t-1} \right) \approx f(y_t | h_t) \cdot \frac{1}{M} \sum_{m=1}^{M} f(h_t | S_t, h_t^{(m)}_{t-1}) \cdot P(S_t | S_{t-1}^{(m)}_{t-1}),$$

where $\theta$ is omitted for simplicity and

$$\ln f(y_t | h_t) = \text{const} - \frac{1}{2} h_t - \frac{y_t^2}{2} \exp(-h_t).$$

Define $\ln f^*(h_t) = - \frac{1}{2} h_t - \frac{y_t^2}{2} \exp(-h_t)$. Applying the first order Taylor expansion to $\ln f^*(h_t)$ around $h_t = \hat{h}_t$ (we set $\hat{h}_t$ to be the smoothed posterior estimates from the Gibbs sampling algorithm) will give the inequality,

$$\ln f^*(h_t) = - \frac{1}{2} h_t - \frac{y_t^2}{2} \exp(-h_t) \leq - \frac{1}{2} h_t - \frac{y_t^2}{2} \exp(-\hat{h}_t) (1 + \hat{h}_t - h_t) \equiv \ln g^*(h_t).$$
We can see that the kernel of the target filtering density \( f(S_t, h_t | Y_t) \) can be bounded as

\[
f^*(h_t) \cdot \frac{1}{M} \sum_{m=1}^{M} f \left( h_t \big| S_{t-1}^{(m)} \right) \cdot P \left( S_t \big| S_{t-1}^{(m)} \right) \leq g^*(h_t) \cdot \frac{1}{M} \sum_{m=1}^{M} f \left( h_t \big| S_{t-1}^{(m)} \right) \cdot P \left( S_t \big| S_{t-1}^{(m)} \right).
\]

The product of \( g^*(h_t) \) and \( f \left( h_t \big| S_{t-1}^{(m)} \right) \cdot P \left( S_t \big| S_{t-1}^{(m)} \right) \) could be expressed as

\[
g^*(h_t) \cdot f \left( h_t \big| S_t, h_{t-1}^{(m)} \right) \cdot P \left( S_t \big| S_{t-1}^{(m)} \right) \propto \pi(S_t, m) \cdot f_N \left( h_t \big| h_t^*(S_t, m), \sigma^2 (1 - \rho^2) \right),
\]

where \( f_N \left( h_t \big| h_t^*(S_t, m), \sigma^2 (1 - \rho^2) \right) \) is the normal density with mean \( h_t^*(S_t, m) \) and variance \( \sigma^2 (1 - \rho^2) \), and

\[
h_t^*(S_t, m) = \mu_0 + \mu_1 S_t + \phi h_{t-1}^{(m)} + \rho \sigma y_{t-1} \exp(-h_{t-1}^{(m)} / 2)
\]

\[
+ \frac{\sigma^2 (1 - \rho^2)}{2} \left[ y_t^2 \exp(-h_t) - 1 \right],
\]

\[
\pi(S_t, m) = \exp \left\{ -\frac{1}{2\sigma^2 (1 - \rho^2)} \left[ \left( \mu_0 + \mu_1 S_t + \phi h_{t-1}^{(m)} \right)^2
\right.
\]

\[
+ \rho \sigma y_{t-1} \exp(-h_{t-1}^{(m)} / 2) \right)^2 - \left( h_t^*(S_t, m) \right)^2 \right\} \cdot P \left( S_t \big| S_{t-1}^{(m)} \right) \).
\]

So, the filtering density \( f(S_t, h_t | Y_t) \) could be written as

\[9\] The derivations of \( h_t^*(S_t, m) \) and \( \pi(S_t, m) \) are provided in the Appendix.
Therefore, we can sample from the filtering density \( f(S_t, h_t|Y_t) \) using the accept-reject algorithm. First, we draw a proposal \( (S_t, h_t) \) from the mixture of \( M \) normal densities

\[
f(S_t, h_t|Y_t) \propto f^*(h_t) \cdot \frac{1}{M} \sum_{m=1}^{M} f\left(h_t|S_t, h_{t-1}^{(m)}\right) \cdot P\left(S_t|S_{t-1}^{(m)}|t-1\right)
\]

\[
\leq g^*(h_t) \cdot \frac{1}{M} \sum_{m=1}^{M} f\left(h_t|S_t, h_{t-1}^{(m)}\right) \cdot P\left(S_t|S_{t-1}^{(m)}|t-1\right)
\]

\[
\propto \frac{1}{M} \sum_{m=1}^{M} \pi(S_t, m) \cdot f_N\left(h_t|h_t^*(S_t, m), \sigma^2(1 - \rho^2)\right).
\]

Therefore, we can sample from the filtering density \( f(S_t, h_t|Y_t) \) using the accept-reject algorithm. First, we draw a proposal \( (S_t, h_t) \) from the mixture of \( M \) normal densities

\[
\sum_{m=1}^{M} \pi^*(S_t, m) \cdot f_N\left(h_t|h_t^*(S_t, m), \sigma^2(1 - \rho^2)\right),
\]

where \( \pi^*(S_t, m) = \pi(S_t, m) / \sum_{m=1}^{M} \pi(S_t, m) \). We could sample from this mixture distribution by first selecting the indices \( (S_t, m) \) with probability \( \pi^*(S_t, m) \), and then sampling from the associated \( f_N\left(h_t|h_t^*(S_t, m), \sigma^2(1 - \rho^2)\right) \). Second, we accept this proposed \( (S_t, h_t) \) with probability \( f^*(h_t)/g^*(h_t) \). If rejected, we return to the first step and draw a new proposed \( (S_t, h_t) \).

By selecting a large \( M \), this filtering sampler will become arbitrarily accurate. We specify \( M = 2500 \) in our studies. Specifically, we initiate the particle filter with
(h_{0|0}^{(m)}, S_{0|0}^{(m)}) \ (m = 1, \cdots, 2500), then run the particle filter sequentially and get
((h_{t|t}^{(m)}, S_{t|t}^{(m)}) \ (m = 1, \cdots, 2500) for each t, t = 1, \cdots, T.

6.1.2 One-step-ahead Prediction Density

For the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model), the one-step-ahead prediction density, \( f(y_{t+1}|Y_t) \), is written as

\[
 f(y_{t+1}|Y_t) = \int f(y_{t+1}|h_{t+1}) \cdot f(h_{t+1}|S_{t+1}, h_t) \cdot P(S_{t+1}|S_t) \\
 \quad \cdot f(S_t, h_t|Y_t) \, dh_t \, ds_{t+1} \, ds_t.
\]

Suppose we have gained \( M \) draws \((h_{t|t}^{(m)}, S_{t|t}^{(m)}) \ (m = 1, \cdots, M)\) sampled from the filtering density \( f(S_t, h_t|Y_t) \) through the particle filter illustrated in section 6.1.1, then, we can sample \( S_{t+1|t}^{(m)} \ (m = 1, \cdots, M)\) using the transition probability \( P(S_{t+1}|S_{t|t}^{(m)}) \) and \( h_{t+1|t}^{(m)} \ (m = 1, \cdots, M)\) from \( f(h_{t+1}|S_{t+1|t}, h_{t|t}^{(m)})\), which is the normal density with mean \( \mu_0 + \mu_1 S_{t+1|t} + \phi h_{t|t}^{(m)} + \rho \sigma y_t \exp(-h_{t|t}^{(m)}/2) \) and variance \( \sigma^2 (1 - \rho^2)\). Using \( h_{t+1|t}^{(m)} \ (m = 1, \cdots, M)\), the one-step-ahead prediction density can be evaluated as

\[
 f(y_{t+1}|Y_t) \approx \frac{1}{M} \sum_{m=1}^{M} f(y_{t+1}|h_{t+1|t}^{(m)}).
\]

where \( f(y_{t+1}|h_{t+1|t}^{(m)}) \sim \mathcal{N}(0, \exp(h_{t+1|t}^{(m)})) \).
This derivation of one-step-ahead prediction will facilitate to calculate diagnostics statistics and evaluate likelihood. In the future sections, we will explain the use of one-step-ahead prediction results.

6.2 Diagnostics of SVLR Model for S&P 500 Index Weekly Return Data

In Econometrics, diagnostic checks are required to assess the adequacy of the model and how well the fitted model accords with the observed data. If the model is invalid, then it can yield false inference, so model checking is crucial to statistical analysis. For different models, the diagnostic methods are different.

Since the SVLR model is nonlinear and non-Gaussian, the residuals from the model are not normal and we could not rely on checking if the standardized residuals are white noise. The diagnostic method for our SVLR model is from Rosenblatt (1952) and Kim et al. (1998), where the filtering densities gained from the particle filter in section 6.1.1 would be used.

After getting $M$ draws $(h_{t|t}^{(m)}, S_{t|t}^{(m)})$ ($m = 1, \cdots, M$) sampled from the filtering density $f(S_t, h_t|Y_t)$, we can sample $S_{t+1|t}^{(m)}$ ($m = 1, \cdots, M$) using the transition probability $P(S_{t+1|t}|S_{t|t}^{(m)})$ and then $h_{t+1|t}^{(m)}$ ($m = 1, \cdots, M$) from $f(h_{t+1|t}|S_{t+1|t}^{(m)}, h_{t|t}^{(m)})$. Let $y_{t+1}^O$ denote the observation of $y_{t+1}$. Given draws $h_{t+1|t}^{(m)}$ ($m = 1, \cdots, M$), it is straightforward to estimate the probability that $y_{t+1}^2$ will be less than $y_{t+1}^O$ conditional on $Y_t$, i.e.
\[ P(y_{t+1}^2 \leq y_{t+1}^o | Y_t) = \int P(y_{t+1}^2 \leq y_{t+1}^o | h_{t+1}) f(h_{t+1} | Y_t) dh_{t+1} \]

\[ \approx \frac{1}{M} \sum_{m=1}^{M} P(y_{t+1}^2 \leq y_{t+1}^o | h_{t+1}^{(m)}) \]

Since the distribution of \( y_{t+1} \) conditional on \( h_{t+1|t}^{(m)} \) is the normal with mean 0 and variance \( \exp(h_{t+1|t}^{(m)}) \), the above probability of \( y_{t+1}^2 \) being less than \( y_{t+1}^o \) conditional on \( h_{t+1|t}^{(m)} \) will be readily available.

We let \( u_{t+1}^M = \frac{1}{M} \sum_{m=1}^{M} P(y_{t+1}^2 \leq y_{t+1}^o | h_{t+1|t}^{(m)}) \). For each \( t = 1, \ldots, T \), under the null hypothesis of a correctly specified model, \( u_{t}^M \) converges in distribution to independently and identically distributed uniform random variables as \( M \to \infty \) (see Rosenblatt (1952) and Kim et al. (1998)). This provides a valid basis for diagnostic checking. These variables can be transformed into the normal distribution. Using the inverse of the normal distribution function \( n_{t}^M = F^{-1}(u_{t}^M) \) gives a sequence of independently and identically standard normally distributed variables, \( n_{1}^M, \ldots, n_{T}^M \). The diagnostic check of the SVLR model can be conducted through testing whether the transformed sequence \( n_{1}^M, \ldots, n_{T}^M \) follows the independent standard normal distribution.

Table 6.1 shows the results of standard diagnostic checks on \( n_{1}^M, \ldots, n_{T}^M \) produced by the fitted SVLR model. Under the correctness of the model, the diagnostics should indicate that the variables are Gaussian white noise. The data set used here is still the
S&P 500 index weekly returns from the first week of April in 1986 to the last week of December in 2011, and we let $M = 2500$. We report the mean, the standard deviation, the skewness, the kurtosis, the overall Jarque–Bera normality statistic and the Ljung-Box test statistic used to test the null hypothesis of no serial correlation for 30 lags. The SVLR model could pass all diagnostic tests at the 5% significance level. Jarque-Bera statistic shows that $n^M_1, \ldots, n^M_T$ could pass the normality test with 95% confidence level. The Ljung-Box statistic shows absence of autocorrelation in $n^M_1, \ldots, n^M_T$ at all lags up to 30 lags. The resulting correlograms and QQ plots are given in Figure 6.1. The graphical diagnostics also show that the SVLR model performs quite well to the data. In the QQ plot, the plots of ordered transformed sequence $n^M_1, \ldots, n^M_T$ against the standard normal quantiles are very close to the 45 degree reference line, which also verify the normality of $n^M_1, \ldots, n^M_T$.

Table 6.1 Diagnostics of the SVLR model

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>S.D.</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>JB</th>
<th>LB (30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVLR Model</td>
<td>0.0491</td>
<td>0.9449</td>
<td>-0.1445</td>
<td>2.9406</td>
<td>4.8822</td>
<td>40.1106</td>
</tr>
</tbody>
</table>

Note: The data used here is still S&P500 index weekly returns.
Figure 6.1 Diagnostic Checks of the SVLR Model

(a) Plots of $n_1^M, \cdots, n_T^M$ produced by the SVLR model.
(b) Autocorrelation of $n^M_1, \ldots, n^M_T$.

(c) Autocorrelation of $y^2_t$. 

6.3 Model Comparison

It is important to check whether the mean of the log-volatility shifts depending on the state and whether there is leverage effect in the volatility. These are equivalent to comparing our SVLR model and the existing SVL or SVR models. Model comparison under a Bayesian framework could be performed through using the posterior odds ratio. Let $Y_T$ denote $(y_1, \ldots, y_T)$, the observation data till time $T$. Then, the posterior odds ratio (abbreviated by POR) between model $i$, $M_i$, and model $j$, $M_j$, is specified by

$$POR = \frac{f(M_i|Y_T)}{f(M_j|Y_T)} = \frac{f(Y_T|M_i)}{f(Y_T|M_j)} \cdot \frac{f(M_i)}{f(M_j)}.$$
where \( f(Y_T|M_i)/f(Y_T|M_j) \) and \( f(M_i)/f(M_j) \) are called Bayes factor and prior odds ratio respectively. Usually the prior odds ratio is set to be one, so the Bayes factor is equal to the POR. The Bayes factor being greater than one means that POR is greater than one, i.e. \( M_i \) is favored over \( M_j \).

In order to evaluate the Bayes factor, we must calculate \( f(Y_T|M_i) \) and \( f(Y_T|M_j) \), the marginal likelihood for two models. Let \( \theta_i \) denote the set of unknown parameters in model \( M_i \). The marginal likelihood of model \( M_i \), by virtue of being the normalizing constant of the posterior density, can be written as

\[
f(Y_T|M_i) = \frac{f(Y_T|M_i, \theta_i) \cdot f(\theta_i|M_i)}{f(\theta_i|M_i, Y_T)},
\]

where \( f(Y_T|M_i, \theta_i) \) is likelihood, \( f(\theta_i|M_i) \) is prior density and \( f(\theta_i|M_i, Y_T) \) is posterior density. The above identity is called basic marginal likelihood identity (BMI) proposed by Chib (1995) and holds for any value of \( \theta_i \). Following Chib (1995), we set \( \theta_i \) equal to its posterior mean \( \theta_i^* \) from the Bayesian Gibbs sampling algorithm. Based on the above expression, an estimate of the marginal likelihood on the log-scale is given by

\[
\log \hat{f}(Y_T|M_i) = \log f(Y_T|M_i, \theta_i^*) + \log f(\theta_i^* |M_i) - \log \hat{f}(\theta_i^* | M_i, Y_T).
\]
It is straightforward to evaluate the prior density $f(\theta^*_i | M_i)$. In the following, we will propose the methods to calculate the likelihood $f(Y_T | M_l, \theta^*_l)$ and estimate the posterior density $f(\theta^*_i | M_b, Y_T)$ for our model, the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model).

### 6.3.1 Likelihood Evaluation for the SVLR Model

The evaluation of the likelihood from the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) will use the results from the Particle Filter and One-step-ahead prediction density illustrated in section 6.1.

For simplicity, we omit $M_l$ and $\theta_l$ in the calculation for the likelihood. Let $Y_t$ denote $(y_1, \cdots, y_t)$, the observation data till time $t$. Then the likelihood can be expressed as

$$f(Y_T) = f(y_1, \cdots, y_T) = f(y_1) \prod_{t=1}^{T-1} f(y_{t+1} | Y_t).$$

For the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model), the one-step-ahead prediction density, $f(y_{t+1} | Y_t)$, is written as

$$f(y_{t+1} | Y_t) = \int f(y_{t+1}, h_{t+1}) \cdot f(h_{t+1} | S_{t+1}, h_t) \cdot P(S_{t+1} | S_t) \cdot f(S_t, h_t | Y_t) dh_{t+1} dh_t dS_{t+1} dS_t.$$

Suppose that we have $M$ draws $(h_{t|t}^{(m)}, S_{t|t}^{(m)})(m = 1, \cdots, M)$ obtained from the filtering density $f(S_t, h_t | Y_t)$ through the particle filter. Then, we can sample $S_{t+1|t}^{(m)}(m = 1, \cdots, M)$
using the transition probability \( P(S_{t+1}|S_t^{(m)}) \) and \( h_{t+1|t}^{(m)} \) (\( m = 1, \cdots, M \)) from
\[
f(h_{t+1|t}^{(m)}, h_{t|t}^{(m)}),
\]
which is the normal density with mean \( \mu_0 + \mu_1 S_{t+1|t}^{(m)} + \phi h_{t|t}^{(m)} + \rho \sigma y_t \exp(-h_{t|t}^{(m)}/2) \) and variance \( \sigma^2 (1 - \rho^2) \). Using \( h_{t+1|t}^{(m)} \) (\( m = 1, \cdots, M \)), the one-step-ahead prediction density for the SVLR model can be evaluated as
\[
f(y_{t+1}|Y_t) \approx \frac{1}{M} \sum_{m=1}^{M} f(y_{t+1}|h_{t+1|t}^{(m)}),
\]
where \( f(y_{t+1}|h_{t+1|t}^{(m)}) \sim \mathcal{N}(0, \exp(h_{t+1|t}^{(m)})) \). Once the one-step-ahead prediction density is available, the likelihood function could be calculated.

### 6.3.2 Estimation of Posterior Densities

After fulfilling the evaluation of the likelihood \( f(Y_T|M_i, \theta^*_i) \), we need to figure out how to estimate the posterior density \( f(\theta^*_i|M_i, Y_T) \) in order to calculate the marginal likelihood for the model \( M_i, f(Y_T|M_i) \). One possible approach is based on the method of kernel smoothing (Ritter and Tanner (1992)). But when \( \theta \) is high dimensional and the model contains latent variables, the estimate will be less accurate. We will adopt Chib (1995)'s posterior density decomposition method, which estimates the posterior density from the output of the Gibbs sampler by exploiting the information in the collection of the complete conditional densities.
The parameters for the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) include the regular parameters \( \theta = (\phi, \sigma, \rho, \mu_0, \mu_1, p, q)' \) and the latent variables \( z = (H, S) \) where \( H = (h_1, \cdots, h_T)' \) and \( S = (S_1, \cdots, S_T)' \). In our Gibbs sampler for the model, we divide \( \theta \) into three blocks, i.e. \( \theta_1 = (\mu_0, \mu_1, \phi), \theta_2 = (\rho, \sigma) \) and \( \theta_3 = (p, q) \) and the complete conditional densities include \( f(\theta_1|Y_T, \theta_2, \theta_3, z), f(\theta_2|Y_T, \theta_1, \theta_3, z), f(\theta_3|Y_T, \theta_1, \theta_2, z) \) and \( f(z|Y_T, \theta_1, \theta_2, \theta_3) \). Our objective now is to estimate the posterior density at the point \( \theta^*, f(\theta^*|Y_T) \), which can be expressed as

\[
f(\theta^*|Y_T) = f(\theta_1^*|Y_T) \cdot f(\theta_2^*|\theta_1^*, Y_T) \cdot f(\theta_3^*|\theta_1^*, \theta_2^*, Y_T),
\]

where we omit the index for model \( M_i \). The first term is the marginal ordinate, while the second and the third terms are both called the reduced conditional ordinate.

The marginal ordinate can be estimated from the draws of the initial Gibbs run, i.e.

\[
\hat{f}(\theta_1^*|Y_T) = G^{-1} \sum_{g=1}^{G} f(\theta_1^*|Y_T, \theta_2^{(g)}, \theta_3^{(g)}, z^{(g)}),
\]

because \( f(\theta_1^*|Y_T) = \int f(\theta_1^*|Y_T, \theta_2, \theta_3, z) \cdot f(\theta_2, \theta_3, z|Y_T) \, d\theta_2 d\theta_3 dz \) and \( \theta_2^{(g)}, \theta_3^{(g)} \) and \( z^{(g)} \) are the draws from the distribution \( \theta_2, \theta_3, z|Y_T \) according to the MCMC theory. Gelfand and Smith (1990) refer to this approach as “Rao-Blackwellization”.

77
But the above technique could not be used to estimate the reduced conditional ordinates. We can take the second term \( f(\theta_2^*|\theta_1^*, Y_T) \) as an example to explain it. We know that 

\[
    f(\theta_2|\theta_1, Y_T) = \int f(\theta_2|\theta_1, \theta_3, z, Y_T) \cdot f(\theta_3, z|\theta_1^*, Y_T) d\theta_3 dz.
\]

However, draws \( \theta_3^{(g)} \) and \( z^{(g)} \) are from the distribution \( \theta_3, z|Y_T \) not from \( \theta_3, z|\theta_1^*, Y_T \), so the complete conditional density of \( \theta_2 \) cannot be averaged directly. The solution is to continue sampling from an additional \( G \)-time iteration with the complete conditional densities

\[
    f(\theta_2|Y_T, \theta_1^*, \theta_3, z), f(\theta_3|Y_T, \theta_1^*, \theta_2, z) \text{ and } f(z|Y_T, \theta_1^*, \theta_2, \theta_3),
\]

where \( \theta_1 \) is fixed at \( \theta_1^* \) in each of the these densities. It can be verified that draws \( \theta_3^{(j)} \) and \( z^{(j)} \) from this run follow the density \( \theta_3, z|\theta_1^*, Y_T \), as required. Consequently,

\[
    \hat{f}(\theta_2^*|\theta_1^*, Y_T) = G^{-1} \sum_{j=1}^{G} f(\theta_2|Y_T, \theta_1^*, \theta_3^{(j)}, z^{(j)})
\]

is the estimate of the reduced conditional ordinate \( f(\theta_2^*|\theta_1^*, Y_T) \). Although this needs an increase in the number of iterations, but it does not require new programming and thus is easy to implement.

Finally, additional \( G \) iterations with the densities

\[
    f(\theta_3|Y_T, \theta_1^*, \theta_2^*, z) \text{ and } f(z|Y_T, \theta_1^*, \theta_2^*, \theta_3)
\]
will produce draws $z^{(j)}$ that follow the distribution $z | \theta_1^*, \theta_2^*, Y_T$. These draws can yield an estimate

$$\hat{f}(\theta_3^* | \theta_1^*, \theta_2^*, Y_T) = G^{-1} \sum_{j=1}^{G} f(\theta_3^* | Y_T, \theta_1^*, \theta_2^*, z^{(j)}).$$

In order to compute the posterior densities by the above approach, it is necessary that all normalizing constants of the full conditional distributions in the Gibbs sampler be known. This requirement is easily satisfied in our model since we all use conjugate priors in our algorithm and the normalizing constants for all the posteriors are available.

After getting the estimates of $\hat{f}(\theta_1^* | Y_T), \hat{f}(\theta_2^* | \theta_1^*, Y_T)$ and $\hat{f}(\theta_3^* | \theta_1^*, \theta_2^*, Y_T)$, the log of the marginal likelihood of the Stochastic Volatility Model with Leverage Effect and Regime Switching (SVLR model) becomes

$$\log \hat{f}(Y_T | M_i) = \log f(Y_T | M_i, \theta^*) + \log \hat{f}(\theta^* | M_i) - \log \hat{f}(\theta_1^* | M_i, Y_T) - \log \hat{f}(\theta_2^* | \theta_1^*, Y_T) - \log \hat{f}(\theta_3^* | \theta_1^*, \theta_2^*, Y_T).$$

### 6.3.3 Comparison between the SVLR model and the existing SV models

In this section we compare the fit of our SVLR model and the existing SV models in the literature, the SVL and the SVR models. The data series used here is still the S&P
500 index weekly returns from the first week of April in 1986 to the last week of December in 2011.

The SVL model is actually a special case for the SVLR model when there is no switching regime and the SVR model is a special case for the SVLR model when there is no leverage effect. The methods for calculating the likelihood and estimating the posterior density of parameters for the SVL and SVR models could be obtained through modifying the methods for the SVLR models. In the procedures for estimating the posterior density, we set the number of MCMC iteration equal to 50,000. In the particle filter to evaluate the likelihood for all the three models, we set the number of simulation $M$ equal to 2500.

Table 6.2 reports the estimates of the log likelihood, the log posterior density and prior densities of the parameters and the log marginal likelihood for all the three models, the SVLR, the SVL model and the SVR model. All the densities are evaluated at the estimated posterior mean of parameters from the MCMC procedures.
### Table 6.2 Comparison between the SVLR, SVL and SVR models under the Bayesian framework

<table>
<thead>
<tr>
<th></th>
<th>SVLR model</th>
<th>SVL model</th>
<th>SVR model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log likelihood</td>
<td>-2889.3</td>
<td>-2910.5</td>
<td>-2907.6</td>
</tr>
<tr>
<td>Log prior density</td>
<td>-6.2884</td>
<td>-8.1555</td>
<td>-4.9133</td>
</tr>
<tr>
<td>Log posterior density</td>
<td>19.5875</td>
<td>10.0359</td>
<td>18.4377</td>
</tr>
<tr>
<td>Log marginal likelihood</td>
<td>-2915.18</td>
<td>-2928.69</td>
<td>-2930.95</td>
</tr>
<tr>
<td>Log Bayes factor of SVLR/SVL</td>
<td>13.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Bayes factor of SVLR/SVR</td>
<td></td>
<td></td>
<td>15.77</td>
</tr>
</tbody>
</table>

The log marginal likelihood of the SVLR model of -2915.18 is larger than ones of the SVL model and SVR model. We could conclude that our SVLR model is favorable over both of the existing SV models in the literature, the SVL model and the SVR model. Incorporating both the leverage effect and the switching regimes into the SV model could better fitting the S&P 500 index weekly returns from the first week of April in 1986 to the last week of December in 2011. These results also show that there really exist the structure change and leverage effect in the volatility process.
7. Bayesian Density Forecasting of SVLR Model

One of the important goals of any econometric analysis is to predict the future values of some variables of interest given the current data and the assumed model. In contrast to a point forecast, a density forecast is an estimate of the probability distribution of the possible future values of the variables of interest and thus provides a complete description of the uncertainty related with a prediction. Recent development in quantitative finance increases the demand for density forecasting. For example, the booming area of financial risk management is dedicated to providing density forecasting of portfolio values and calculating Value-at-Risk (VaR). The framework of Bayesian inference analysis produces a convenient way to conduct the density forecasting.

Under the Bayesian framework, the problem of density forecasting is solved by calculating the Bayesian prediction density, which is defined as the distribution of future observation values $y_f$ conditional on the current data $Y = (y_1, \cdots, y_T)$ and the assumed model $M$, but marginalized over the parameters $\theta$. The specific formulation for the prediction density is

$$ f(y_f|Y, M) = \int f(y_f|Y, M, \theta)\pi(\theta|Y, M)d\theta, $$

where the prediction density $f(y_f|Y, M)$ is the integral of the conditional density $f(y_f|Y, M, \theta)$ with respect to the posterior density $\pi(\theta|Y, M)$ of $\theta$. In general, this integral, i.e. the prediction density, is not available in closed form. However, the MCMC outputs, a sample of (correlated) draws
\[ \theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(M)} \sim \pi(\theta | Y, M), \]

could help obtain the approximation of the prediction density. We can use the posterior draws to produce a sample of draws from the prediction density. Specifically, it is done by appending a step at the end of the MCMC iterations where for each value \( \theta^{(j)} \) we simulates

\[ y_f^{(j)} \sim f(y_f | Y, M, \theta^{(j)}) , \quad j \leq M, \]

from the density of the future observations, conditional on \( \theta^{(j)} \). Then the collection of the simulated draws \( \{ y_f^{(1)}, \ldots, y_f^{(M)} \} \) could be regarded as the samples from the Bayesian prediction density \( f(y_f | Y, M) \). In this chapter, we will predict the future observations of stock returns through sampling from the Bayesian prediction density of SVLR model and use the S&P 500 index data to evaluate the density forecasting capability of the SVLR model.

7.1 The Bayesian Prediction Density for the SVLR Model

The SVLR model has two latent variables, the unobserved volatility \( h_t \) and the regime state variable \( S_t \). In order to sample the prediction density for this latent structure, we need to do some further modification to the procedure we mentioned above.
7.1.1 The Bayesian Prediction Density for the Latent-Structured Model

Suppose that \( z_f \) denotes the latent data in the prediction period and \( z \) represents the latent data in the sample period. Let \( \psi = (\theta, z) \) and suppose that the MCMC sampler in chapter 3 produces the sampled draws

\[
\psi^{(1)}, \ldots, \psi^{(M)} \sim \pi(\psi | \mathcal{Y}, \mathcal{M}).
\]

In this situation, the prediction density can be expressed as

\[
f(y_f | \mathcal{Y}, \mathcal{M}) = \int f(y_f | \mathcal{Y}, \mathcal{M}, z_f, \psi) \pi(z_f | \mathcal{Y}, \mathcal{M}, \psi) \pi(\psi | \mathcal{Y}, \mathcal{M}) \, dz_f \, d\psi,
\]

which could be again sampled through utilizing the MCMC outputs, i.e. the (correlated) samples from the posterior density, \( \psi^{(j)} \sim \pi(\psi | \mathcal{Y}, \mathcal{M}) \). At the end of the MCMC iterations, we append a step to simulate

\[
z_f^{(j)} \sim \pi(z_f | \mathcal{Y}, \mathcal{M}, \psi^{(j)}), \quad y_f^{(j)} \sim f\left(y_f | \mathcal{Y}, \mathcal{M}, z_f^{(j)}, \psi^{(j)}\right), \quad j \leq M.
\]

The collection of the simulated values of \( \{y_f^{(1)}, \ldots, y_f^{(M)}\} \) could be regarded as the samples from the Bayesian prediction density for the model with latent data structure. So, we could utilize this two-step process to do Bayesian density forecasting for our SVLR model.
7.1.2 The Bayesian Density Forecasting of the SVLR Model

Since our SVLR model has unobservable latent variables, the volatility \( h_t \) and the regime state variable \( S_t \), we could utilize the method introduced in section 7.1.1 to get the prediction density of the variables we are interested in.

In this section, we use our SVLR model to do the density forecasting of the log return of the S&P 500 index data. Suppose that we are interested in predicting the one-step-ahead prediction of the log return \( y_f = y_{T+1} \), given the available data information through time \( T, Y = (y_1, \ldots, y_T) \). And suppose that the Gibbs sampler in Chapter 3 has been used to deliver draws on \( \theta = (\phi, \sigma, \rho, \mu_0, \mu_1, p, q)' \), \( H = (h_1, \ldots, h_T)' \) and \( S = (S_1, \ldots, S_T)' \). In order to predict \( y_{T+1} \), we take each draw from the sampler and sample

\[
S_{T+1}^{(j)} \sim p \left( S_{T+1} | S_T^{(j)}, p^{(j)}, q^{(j)} \right), \quad j \leq M
\]

from the Markov chain (a two point discrete distribution), and then sample

\[
h_{T+1}^{(j)} \sim \mathcal{N} \left( \mu_0^{(j)} + \mu_1^{(j)} S_{T+1}^{(j)} + \phi^{(j)} h_T^{(j)} + \rho^{(j)} \sigma^{(j)} y_T \exp \left( -\frac{h_T^{(j)}}{2} \right), \sigma^2 (j) \left( 1 - \rho^2 (j) \right) \right), \quad j \leq M
\]

from the state equation of the SVLR model, and finally sample
If one is interested in predicting $y_{T+2}$, the sample $y^{(j)}_{T+2}$ could be drawn in the same way after $S_{T+2}^{(j)}$ and $h_{T+2}^{(j)}$ are simulated from the Markov chain $p(S_{T+2}^{(j)}|S_{T+1}^{(j)},p^{(j)},q^{(j)})$ and the volatility process $N\left(\mu_{0}^{(j)} + \mu_{1}^{(j)}S_{T+2}^{(j)} + \phi^{(j)}h_{T+1}^{(j)} + \rho^{(j)}\sigma^{(j)}y_{T+1}^{(j)} \exp\left(-\frac{h_{T+1}^{(j)}}{2}\right), \sigma^{2}(j)(1 - \rho^{2}(j))\right)$. These steps could be iterated for any number of periods into the future and the whole process repeated for each simulated value for the parameters and the latent variables.

Here, we will do a one-step-ahead Bayesian density forecasting of SVLR model. We choose the weekly S&P index return data from 04/07/1986 through 03/07/2011 as the historically available data and then do the one-step-ahead prediction. So, the in-sample data size for our study is 1300 and we will use the information from the in-sample data to do the one-step-ahead prediction of the S&P 500 index weekly return on 03/14/2011. Figure 7.1 shows the histogram of sampled draws from the Bayesian prediction density of $f(y_{T+1} | Y_T)$, where $T=1300$ and $Y_T$ denote the in-sample data. We superimpose the fitted normal density (red line) with the parameters estimated from the drawn samples in the same graph. It could be seen that the forecast density is not normal. The skewness of the forecast density is -0.1316 and the kurtosis is 5.1139. The result of the Jarque-Bera test
is to reject the normality of the drawn samples from the Bayesian prediction density of $f(y_{T+1}|Y_T)$ at the 5% significance level.

**Figure 7.1 The Histogram Of Samples From The Bayesian Prediction Density**

![Histogram of Samples](image)

**7.2 The Evaluation of the Bayesian Density Forecasting of SVLR Model**

Actually in literature, less attention has been given to evaluating the density forecasts. We will firstly introduce the concept of loss function, study the relationships between density forecasts and loss functions and get the implications for density forecasts.
evaluation. In the following, we will introduce the probability integral transform (PIT) method proposed by Diebold, Gunther and Tay (1998). In the framework of PIT method, we will do an empirical study with S&P 500 weekly return series to evaluate the density forecast of our SVLR model.

7.2.1 Loss Functions and Density Forecasts

It is argued that density forecasts could also be evaluated in a decision context. Specifically, the forecast user has a loss function $L(a(p(y)), y)$ which depends on the action $a$ chosen by the forecast user on the basis of the density forecast $p(y)$ and on the realization of the forecast variable $y$. The forecast user would choose an action $a^*$ to minimize expected loss calculated as if the density forecast $p(y)$ is correct, i.e.

$$a^*(p(y)) = \arg\min_{a \in A} \int L(a, y)p(y)dy,$$

where $A$ denotes all possible action choices for the forecast user. The choice $a^*$ will incur loss $L(a^*, y)$, which is a random variable whose expected value with respect to the true density $f(y)$ is

$$E[L(a^*, y)] = \int L(a^*, y)f(y)dy.$$
Obviously different density forecasts will lead to different action choices and hence different expected loss. The better a density forecast is, the lower its expected loss. Simply speaking, the user will prefer forecast \( p_j(y) \) to forecast \( p_k(y) \) if

\[
E[L(a_j^*, y)] < E[L(a_k^*, y)], \text{ i.e. } \int L(a_j^*, y)f(y)dy < \int L(a_k^*, y)f(y)dy.
\]

Forecast users determine their own loss functions subjectively. It is impossible to find a ranking of forecasts with which all users agree, since different loss functions may lead to totally different rankings for the same sets of density forecasts.

Diebold, Gunther and Tay (1998) propose that the feasible way of evaluating density forecasts could be seeing if the density forecast \( p(y) \) coincides with the true density \( f(y) \). If some density forecast coincides with the true density, then it will be preferred to all forecast users, regardless of loss functions. Formally speaking, if \( p_j(y) = f(y) \), then according to forecast users’ minimization principle,

\[
\int L(a_j^*, y)p_j(y)dy < \int L(a_k^*, y)p_k(y)dy, \forall k
\]

i.e. \( \int L(a_j^*, y)f(y)dy < \int L(a_k^*, y)f(y)dy, \forall k. \)

That means \( a_j^* \) minimize expected loss with respect to the true density over all possible actions and density forecast \( p_j(y) \) will be preferred to all other forecasts, regardless of loss functions.
So, we could evaluate density forecasts through assessing whether the forecast densities are ‘correct’ (close to true densities enough), that is, whether 

\[
\{p_t(y_t|\Omega_t)\}_{t=1}^n = \{f_t(y_t|\Omega_t)\}_{t=1}^n.
\]

Here \(\{f_t(y_t|\Omega_t)\}_{t=1}^n\) is the sequence of conditional true densities for a series of \(y_t\), where \(\Omega_t = \{y_{t-1}, y_{t-2}, \ldots\}\), and \(\{p_t(y_t|\Omega_t)\}_{t=1}^n\) is the corresponding sequence of one-step-ahead density forecasts. The task of checking whether \(\{p_t(y_t|\Omega_t)\}_{t=1}^n = \{f_t(y_t|\Omega_t)\}_{t=1}^n\) is not easy, since we could not know the form of \(f_t(y_t|\Omega_t)\). The probability integral transform (PIT) method is used to solve for this problem.

7.2.2 The Probability Integral Transform (PIT)

The probability integral transform (PIT) can date back to Rosenblatt (1952). Diebold, Gunther and Tay (1998) adopt PIT to do density forecasts evaluations. The formal explanation of this method is that for a sample of \(n\) one-step-ahead forecasts \(\{p_t(y_t|\Omega_t)\}_{t=1}^n\) and the corresponding realizations of \(\{y_t\}_{t=1}^n\), the probability integral transform of the realized variables \(y_t\) with respect to the forecast densities is

\[
    z_t = \int_{-\infty}^{y_t} p_t(u)du = P_t(y_t), \quad t = 1, \ldots, n,
\]

where \(P_t(y_t)\) is the cumulative density function corresponding to the forecast density \(p_t(y_t)\) evaluated at \(y_t\). If \(P_t(\cdot)\) is ‘correct’, then the transformed series, \(\{z_t\}_{t=1}^n\), should be i.i.d. \(U(0, 1)\).
Let’s firstly check the density of $z_t$,

$$
q_t(z_t) = f_t(y_t) \cdot \left| \frac{\partial y_t}{\partial z_t} \right| = f_t(P_t^{-1}(z_t)) \cdot \left| \frac{\partial P_t^{-1}(z_t)}{\partial z_t} \right|.
$$

Since $p_t(y_t) = \partial P_t(y_t)/\partial y_t$ and $y_t = P_t^{-1}(z_t)$, so

$$
p_t(P_t^{-1}(z_t)) = \partial P_t(P_t^{-1}(z_t))/\partial P_t^{-1}(z_t) = \partial z_t/\partial P_t^{-1}(z_t) \text{ and then}
$$

$$
q_t(z_t) = f_t(P_t^{-1}(z_t)) \cdot \left| \frac{\partial P_t^{-1}(z_t)}{\partial z_t} \right| = \frac{f_t(P_t^{-1}(z_t))}{p_t(P_t^{-1}(z_t))}.
$$

If $p_t(y_t) = f_t(y_t)$, then $q_t(z_t)$ is simply the $U(0, 1)$ density. Beyond the one-period characterization for the density of $z$, when we consider the entire $z$ sequence $\{z_t\}_{t=1}^n$, they should be $i.i.d. U(0, 1)$. The independence property is obvious in the case of iid forecasts and also extends to the case of time dependent density forecasts provided that the forecasts are based on an information set that contains the past history of the forecast variable (Diebold, Gunther, Tay, 1998).

### 7.2.3 The Evaluation of Bayesian Density Forecasts of SVLR Model

In this section we will study the density forecast of weekly S&P 500 index return from 04/07/1986 through 03/12/2012. We split the sample in to in-sample (04/07/1986 –
03/07/2011) and out-of-sample (03/14/2011 – 03/12/2012) periods for model estimation and density forecast evaluation. The in-sample data size is 1300 and the out-of-sample data size is 53, nearly one-year period.

We firstly use the in-sample data from 04/07/1986 through 03/07/2011 to do the estimation through the Bayesian MCMC method based on the Gibbs sampler algorithm introduced in Chapter 3. And then we could obtain the density forecast of the one-step-ahead return variable through sampling from the Bayesian prediction density of SVLR model described in section 7.1.2 and calculate the probability integral transformed variable. After that, we increase the in-sample data size through adding one out-of-sample data into the in-sample data set, do re-estimation and density forecasting and calculate the probability integral transformed variable once again. We will iterate the same steps until getting the last probability integral transformed variable for the return on 03/12/2012. In order to evaluate whether the SVLR model is adequate for the out-of-sample forecasting of the weekly S&P 500 index returns, we need to test whether the transformed series, \( \{z_t\}_{t=1}^{53} \), are independent \( U(0, 1) \) variates.

Firstly, we could employ the graphical methods to evaluate if the probability integral transformed series \( \{z_t\}_{t=1}^{53} \) are \( i.i.d. \). Diebold, Gunther and Tay (1998) suggest using the visualization tools, the correlogram supplemented with the Bartlett confidence interval, to assess it. For example, the serial correlations in the z serie indicate that the density forecast could not capture the conditional mean dynamics adequately. Since we are also interested in the possible nonlinear forms of dependence in the return, we examine not
only the correlograms of \((z - \bar{z})\), but also those of the powers of \((z - \bar{z})\). We will examine the correlograms of \((z - \bar{z}), (z - \bar{z})^2, (z - \bar{z})^3\) and \((z - \bar{z})^4\). They will reveal whether the Bayesian density forecast of SVLR model is adequate to describe the dynamics of weekly S&P 500 index returns in conditional mean, conditional variance, conditional skewness and conditional kurtosis. In Figure 7.2, we show the correlograms of \((z - \bar{z}), (z - \bar{z})^2, (z - \bar{z})^3\) and \((z - \bar{z})^4\). The sample autocorrelation functions fail to show the evidence of dependence existing in the probability integral transformed series and we could conclude that they are independent series.

**Figure 7.2 The Autocorrelation Functions of Powers of \(z\)**

(a) Sample autocorrelations of \((z - \bar{z})\)  
(b) Sample autocorrelations of \((z - \bar{z})^2\)
The uniformity of the probability integral transformed series could be tested by plotting the empirical cumulative distribution function (blue line) and comparing it with a 45°-degree line (red line) which is the cumulative distribution function of the uniform distribution. The test results could be found in Figure 7.3. In addition, the Kolmogorov-Smirnov test (K-S test) could be employed to test the null hypothesis of the probability integral transformed series \( \{z_t\}_{t=1}^{53} \) being from the i.i.d. \( U(0, 1) \). The Kolmogorov-Smirnov statistic is the maximum distance between the empirical cumulative distribution function of the probability integral transformed \( z \) and the cumulative distribution function of \( U(0, 1) \). The K-S test result does not reject the null hypothesis of the PIT series \( z \) being from the i.i.d. \( U(0, 1) \) at the 5% significance level.
Up to here, we could conclude that the out-of-sample forecasting performance of the SVLR model for the weekly S&P 500 index return data is good enough. The SVLR model would be a potentially successful model in forecasting the asset returns in Finance, which would be also one of the important applications of the SVLR model in financial studies.
8. Stock Market Volatility and Business Cycle

In this chapter, we will investigate what kind of connection exists between stock market volatility and the business cycle from a new perspective. The volatility processes of stock returns are characterized by periods of high-volatilities separated by periods of low-volatilities. On the other side, the output growth process in real economy is characterized by periods of low-leveled growth separated by periods of high-leveled growth. We are interested in figuring out whether these two processes are influenced by two related driven forces. In order to answer this question, we will study our Stochastic Volatility with Leverage Effect and Regime Switching Model (SVLR model) and the Regime Switching Business Cycles Model proposed by Hamilton (1989) jointly. We will study whether the stock market volatility process and the real output process are driven by two related unobserved variables. First, we introduce the industrial production model to describe the output growth process for the real economy. Second, we design a bivariates framework in which we combine our SVLR model with the industrial production model to study the connection between stock market and industrial production from a new perspective. Third, we employ this framework in the empirical environment to figure out if it could improve identifying and forecasting business cycles.

8.1 Industrial Production Model

Hamilton (1989) proposes a new approach to the business cycle, where the real output growth may follow a regime switching autoregressive process with the regime being associated with economic recessions. Specifically, if the economy is expanding, the mean
level of the autoregressive process is high, vice versa. We will utilize this idea to do our studies to stock market volatility and business cycles.

The industrial production model we will use takes the simplified form

\[ y_t^+ = m_0 + m_1 S_t^+ + e_t \]  
(8.1)

where \( e_t = \alpha e_{t-1} + \epsilon_t \).  
(8.2)

We let \( y_t^+ \) denote the real output growth rate. \( S_t^+ \) is an unobserved latent variable that reflects the state of business cycle. \( S_t^+ = 0 \) indicates that the economy is in expansion and \( S_t^+ = 1 \) means that the economy is in recession. When \( S_t^+ = 0 \), the average growth rate of industrial production is given by \( m_0 \) whereas when \( S_t^+ = 1 \), the average growth rate of industrial production is \( (m_0 + m_1) \). The deviation of output growth rate from the regime-specific mean level follows an AR (1) process whose innovation \( \epsilon_t \) is assumed to be \( i.i.d. \) \( \mathcal{N}(0, \sigma_\varepsilon^2) \). Here, the business cycle regime is also the outcome of a two-state Markov chain with the transition probability as

\[ P(S_t^+ = 0 \mid S_{t-1}^+ = 0) = q^+ \text{ and } P(S_t^+ = 1 \mid S_{t-1}^+ = 1) = p^+. \]

In a word, the above equations model a change in the business cycle phase as a shift in the average growth rate. The parameters vectors in this model include the regular vector \( \theta^+ = (m_0, m_1, \alpha, \sigma_\varepsilon^2, p^+, q^+) \)' and the latent vector \( \mathbf{S}^+ = (S_1^+, \cdots, S_T^+) \)'.
for the identification about the regimes, we constrain $m_1 < 0$. That is, regime 1 represents the economic recession.

In order to estimate the industrial production model, we also use the Bayesian MCMC method based on the Gibbs sampler. In this section, we just introduce how to generate the samples from the posterior conditional densities of vector $(m_0, m_1)'$ and $(S_t^+, \cdots, S_T^+)'$. The methods for generating the samples from the posterior conditional densities of $\alpha$, $\sigma_\epsilon^2$ and $(p^+, q^+)'$ could be inferred from the algorithm in Chapter 3.

To draw the samples from the conditional posterior density, $f(m_0, m_1 | \alpha, \sigma_\epsilon^2, S^+, Y^+)$, we need to rewrite equations (8.1) as

$$y_t^+ = m_0 + m_1 S_t^+ + e_t \quad \text{i.e.} \quad e_t = y_t^+ - (m_0 + m_1 S_t^+). \quad (8.3)$$

Then putting equation (8.3) into equation (8.2) yields that

$$y_t^+ - (m_0 + m_1 S_t^+) = \alpha[y_{t-1}^+ - (m_0 + m_1 S_{t-1}^+)] + \epsilon_t$$

$$i.e. \ y_t^+ - \alpha y_{t-1}^+ = m_0(1 - \alpha) + m_1(S_t^+ - \alpha S_{t-1}^+) + \epsilon_t. \quad (8.4)$$

Dividing both sides of equation (8.4) by $\sigma_\epsilon$ let us get

$$\hat{y}_t = m_0 x_{0t} + m_1 x_{1t} + u_t, \quad u_t \sim i.i.d. N(0, 1), t = 2, \cdots, T, \quad (8.7)$$

where $\hat{y}_t = \frac{y_t^+ - \alpha y_{t-1}^+}{\sigma_\epsilon}$, $x_{0t} = \frac{1 - \alpha}{\sigma_\epsilon}$, $x_{1t} = \frac{S_t^+ - \alpha S_{t-1}^+}{\sigma_\epsilon}$.
After writing equation (8.7) in matrix notation, we have

\[ \tilde{Y} = X\tilde{m} + V, \quad V \sim \mathcal{N}(0, I_{T-1}), \quad \text{(8.8)} \]

where \( \tilde{m} = (m_0, m_1)' \). We assume that \( \tilde{m} \) has a normal prior density as \( \tilde{m} \sim \mathcal{N}(\mathbf{d}_0, \mathbf{D}_0) \), where \( \mathbf{d}_0 \) and \( \mathbf{D}_0 \) are known. Then the conditional posterior distribution for \( \tilde{m} \) becomes \( \tilde{m}|\theta^+_{-\tilde{m}}, S^+, Y^+ \sim \mathcal{N}(d_1, D_1) \), where \( d_1 = (D_0^{-1} + X'X)^{-1}(D_0^{-1}d_0 + X'\tilde{Y}), D_1 = (D_0^{-1} + X'X)^{-1} \). We first draw \( \tilde{m} = (m_0, m_1)' \) from the multivariate posterior distribution. To constrain \( m_1 < 0 \), if the generated value of \( m_1 \) is larger than 0, we discard the draws. Otherwise, we save them.

We can also use the multi-move sampler to draw the sequence of \( \{S_t^+\} \), i.e. generating the samples from the joint distribution \( f(S_1^+, \cdots, S_T^+|\theta^+, Y^+) \) at once. In this step, we need to use the Hamilton filter and the backward sampling methods. We know the industrial production could be specified as

\[ y_t^+ - (m_0 + m_1 S_t^+) = \alpha[y_{t-1}^+ - (m_0 + m_1 S_{t-1}^+)] + \epsilon_t, \]

so the relevant state of the system at date \( t \) could be completely summarized by \( S_t^+ \) and \( S_{t-1}^+ \). We here construct a latent variable \( \zeta_t \), thus the industrial production model could be described as a process with \( \zeta_t \) following a 4-state Markov chain, that is
\[ \zeta_t = 1, \text{ if } S^+_t = 0 \text{ and } S^+_{t-1} = 0 \]
\[ \zeta_t = 2, \text{ if } S^+_t = 0 \text{ and } S^+_{t-1} = 1 \]
\[ \zeta_t = 3, \text{ if } S^+_t = 1 \text{ and } S^+_{t-1} = 0 \]
\[ \zeta_t = 4, \text{ if } S^+_t = 1 \text{ and } S^+_{t-1} = 1, \]

and the 4-state Markov chain has the transition probability matrix as

\[
P_\zeta = \begin{bmatrix}
q^+ & q^+ & 0 & 0 \\
0 & 0 & 1-p^+ & 1-p^+ \\
1-q^+ & 1-q^+ & 0 & 0 \\
0 & 0 & p^+ & p^+ 
\end{bmatrix}.
\]

The Hamilton filter could provide

\[
f(\zeta_t = j|\hat{Y}^+_t, \theta^+) = \frac{f(\zeta_t = j|\hat{Y}^+_{t-1}, \theta^+) \cdot f(y^+_t|\zeta_t = j, \hat{Y}^+_{t-1}, \theta^+)}{\sum_{j=1}^4 f(\zeta_t = j|\hat{Y}^+_{t-1}, \theta^+) \cdot f(y^+_t|\zeta_t = j, \hat{Y}^+_{t-1}, \theta^+)}
\]

and

\[
f(\zeta_{t+1} = j|\hat{Y}^+_t, \theta^+) = \sum_{i=1}^4 P_{ij,\zeta} \cdot f(\zeta_t = i|\hat{Y}^+_t, \theta^+),
\]

where \( \hat{Y}^+_t \) represents the data information for real output growth rates through time \( t \). Then, the filtered probability for state variable \( S^+_t \) could be calculated as

\[
f(S^+_t = 0|\hat{Y}^+_t, \theta^+) = f(\zeta_t = 1|\hat{Y}^+_t, \theta^+) + f(\zeta_t = 2|\hat{Y}^+_t, \theta^+)
\]
\[
f(S^+_t = 1|\hat{Y}^+_t, \theta^+) = f(\zeta_t = 3|\hat{Y}^+_t, \theta^+) + f(\zeta_t = 4|\hat{Y}^+_t, \theta^+).
\]

(8.9)
In order to generated the samples of \( \{ S_t^+ \} \) from \( f(S_t^+, \cdots, S_T^+ | \theta^+, Y^+ ) \) at once, the last draw of \( S_T^+ \) could be generated directly from \( f(S_T^+ | \bar{Y}_T^+, \theta^+) \) and the draws of \( S_t^+ \) for \( t = T - 1, \cdots, 1 \) could be generated from

\[
f(S_t^+ | S_{t+1}^+, \bar{Y}_t^+, \theta^+) = \frac{f(S_t^+, S_{t+1}^+ | \bar{Y}_t^+, \theta^+)}{f(S_{t+1}^+ | \bar{Y}_t^+, \theta^+)}
= \frac{f(S_{t+1}^+ | S_t^+, \bar{Y}_t^+, \theta^+)}{\sum_{k=0}^1 f(S_{t+1}^+ = k, \bar{Y}_t^+, \theta^+) \cdot f(S_t^+ = k | \bar{Y}_t^+, \theta^+)}
\]

where \( f(S_t^+ | \bar{Y}_t^+, \theta^+) \) is available from equation (8.9) and \( f(S_{t+1}^+ | S_t^+, \bar{Y}_t^+, \theta^+) \) is the transition probability. Since we know that

\[
f(S_1^+, \cdots, S_T^+ | \theta^+, Y^+) = f(S_T^+ | \bar{Y}_T^+, \theta^+) \cdot \prod_{t=1}^{T-1} f(S_t^+ | S_{t+1}^+, \bar{Y}_t^+, \theta^+),
\]

these generated samples from \( f(S_T^+ | \bar{Y}_T^+, \theta^+) \) and \( f(S_{t+1}^+ | S_t^+, \bar{Y}_t^+, \theta^+) \) for \( t = T - 1, \cdots, 1 \) could be regarded as the generated samples from \( f(S_1^+, \cdots, S_T^+ | \theta^+, Y^+) \) at once.

### 8.2 Stock Market Volatility and Business Cycles

In this section, our task is to figure out if there is any link between the stock market volatility process described in our SVLR Model and the industrial production process described in the Industrial Production Model introduced in the above section. In order to answer this question, we will try to study the connection between the phase of stock
market volatility $S_t$ and the phase of the industrial production $S_t^+$. Here, we hypothesize three possibilities to study. One hypothesis is that $S_t$ is completely unrelated to $S_t^+$, i.e. shifts in stock market volatility are completely independently with economic forces driving the business cycle for all $t$ and $\tau$. We call this possibility as Model A. A second hypothesis, Model B, suggests that the factor that causes a shift in stock market volatility is exactly the same factor that drives the economy to go into a recession. i.e. $S_t=S_t^+$. The third possibility, Model C, is that stock market volatility process and the industrial production process are driven by the same fundamental economic forces but are not in the same phase. Specifically, the response of stock market to the recession could be earlier than the falling in industrial production related to the recession. This makes sense since usually the participants in the stock market are forward-looking and an incipient recession may affect the stock market before industrial production starts to fall. In order to express this possibility, we impose $S_{t-1}=S_t^+$. Model B and Model C allow us study the stock market volatility and business cycle jointly and are new ways to uncover the dependence between these two processes. Schwert (1989) does regression of measures of stock volatility on a dummy that takes the value of unity when the economy is in recession and gets the evidence that when the economy is in recession the stock market volatility is higher than the one when the economy is in expansion. One problem of this method is that there may exist error for the measurement of the stock market volatility. Another problem from the method by Schwert (1989) is that the value of dummy variable is available only after NBER release the dates at which the recessions begin and end. The joint specifications in Model B and Model C have the potential that the stock market
volatility would be useful to identify and forecast the business cycles. The Bayesian MCMC methods will be also used in the estimations of Model B and Model C.

The steps in the algorithm for estimating the regular parameters and the latent stock volatilities and state variables are similar to the ones we introduced for SVLR model and Industrial Production model. However under the bi-variate frameworks of Model B and Model C, estimations of the latent state variables, the one for stock market volatility $S_t$ or the one for industrial production process $S_t^+$, need to be explained further. Another point needed to mention is we restrict the auto-coefficient in the stock volatility process like $0 < \phi < 1$ to do more efficient inference. The reasons for it are: (1) the existing literatures all get the positive estimates for $\phi$; (2) the volatility clustering feature in financial data conforms to the fact that high volatility is usually followed by next period's low volatility.

8.2.1 Bivariate Framework – Model B

In this section, we illustrate how to estimate the latent variables of $S_t$ and $S_t^+$ in Model B. Since we have imposed that $S_t=S_t^+$ under Model B, whereas the state of stock market volatility at time $t$ is summarized by $S_t$ and the state of industrial production process is summarized by $S_t^+$ and $S_{t-1}^+$, then we just need to know $S_t$ and $S_{t-1}$ in order to summarize the state for both stock market and the industrial production. We could define a new latent variable, $u_t$, which summarize both the state of stock market and the state of industrial production,
This new latent variable actually follows a 4-state Markov process with the transition matrix as

\[
P = \begin{bmatrix}
    q & q & 0 & 0 \\
    0 & 0 & 1-p & 1-p \\
    1-q & 1-q & 0 & 0 \\
    0 & 0 & p & p
\end{bmatrix},
\]

where \( P_{ji} = \text{Prob}(t_t = j | t_{t-1} = i) \). Then the probability of the stock market being in a high-volatility state or the industrial production process being in a low-growth regime becomes \( P(S_t = 0) = P(t_t = 1) + P(t_t = 2) \), vice versa. In the following, we will explain the step in the Gibbs sampler for Model B where we use the multi-move sampler to draw the sequence of \( \{S_t\} \), i.e. generate the samples from the joint distribution \( f(S_1, \cdots, S_T | \mathbf{H}, \mathbf{Y}^+) \), with \( \mathbf{H} \) denoting the stock volatilities all through time \( T \), \( \mathbf{Y}^+ \) representing all the data for the industrial production and the regular parameters being not indicated here for the notation convenience. Through the Hamilton Filter, we could get

\[
P(t_t = j | \bar{H}_t, \bar{Y}_t^+) = \frac{g(t_t = j | \bar{H}_{t-1}, \bar{Y}_{t-1}^+) \cdot \Pr(h_t, y_t^+ | t_t = j, \bar{H}_{t-1}, \bar{Y}_{t-1}^+)}{\sum_{j=1}^4 g(t_t = j | \bar{H}_{t-1}, \bar{Y}_{t-1}^+) \cdot \Pr(h_t, y_t^+ | t_t = j, \bar{H}_{t-1}, \bar{Y}_{t-1}^+)}
\]
\[ P(t_{t+1} = j | \tilde{H}_t, \tilde{Y}_t^+) = \sum_{i=1}^{4} p_{ji} \cdot P(t_t = j | \tilde{H}_t, \tilde{Y}_t^+), \]

\[ i, j = 1, 2, 3, 4 \text{ and } t = 1, \ldots, T. \]

The important point here is that \( \Pr(h_t, y_t^+ | t_t = j, \tilde{H}_{t-1}, \tilde{Y}_{t-1}^+) \) is the joint density involving the data from both the stock market volatility and industrial production. For example,

\[ \Pr(h_t, y_t^+ | t_t = 1, \tilde{H}_{t-1}, \tilde{Y}_{t-1}^+) = f_N(h_t | \mu_0 + \phi h_{t-1}, \sigma^2) \cdot f_N(y_t^+ | \alpha y_{t-1}^+ + m_0(1 - \alpha), \sigma^2), \]

where \( f_N(\cdot | \mu, \sigma^2) \) means the normal density with mean \( \mu \) and variance \( \sigma^2 \). After setting the initialized values, we could iterate the Hamilton Filter and get the filtered density of the state variable \( t_t \) for each period time. Then we could get the filtered density of variable \( S_t \),

\[ P(S_t = 0 | \tilde{H}_t, \tilde{Y}_t^+) = P(t_t = 1 | \tilde{H}_t, \tilde{Y}_t^+) + P(t_t = 2 | \tilde{H}_t, \tilde{Y}_t^+) \]
\[ P(S_t = 1 | \tilde{H}_t, \tilde{Y}_t^+) = P(t_t = 3 | \tilde{H}_t, \tilde{Y}_t^+) + P(t_t = 4 | \tilde{H}_t, \tilde{Y}_t^+). \]

We could still use the Backward Sampling and Forward Filtering (BSFF) method to generated the samples from \( f(S_1, \ldots, S_T | H, Y^+) \) at once, which is a critical step to complete the Gibbs sampler for Model B. The sample of \( S_T \) could be firstly generated.
from \( P(S_T|\tilde{H}_T, \tilde{Y}_T^+) \), and then the samples of \( S_t \), for \( t = T - 1, \ldots, 1 \), could be generated recursively from

\[
P(S_t|S_{t+1}, \tilde{H}_t, \tilde{Y}_t^+) = \frac{P(S_{t+1}|S_t) \cdot P(S_t|\tilde{H}_t, \tilde{Y}_t^+)}{\sum_{j=0}^{1} P(S_{t+1}|S_t = j) \cdot P(S_t = j|\tilde{H}_t, \tilde{Y}_t^+)}.
\]

### 8.2.2 Bivariate Framework – Model C

Model C specifies that \( S_{t-1} = S_t^+ \), i.e. the state of stock market volatility and the state of industrial production are driven by the same economic force but the response in stock markets are one-period earlier than in the industrial production process. Since we need the state variables of \( S_t \) to describe the stock market volatility process and \( S_t^+ \) and \( S_{t-1}^+ \) to describe the industrial production process, then we need to use \( S_t, S_{t-1} \) and \( S_{t-2} \) to summarize the current phase for both stock market volatility and industrial productions.

We could suppose a latent variable \( L_t \) which takes on one of the 8 integer values (1, 2, 3, 4, 5, 6, 7, 8), i.e.

\[
L_t = \begin{cases} 
1 & \text{if } S_t = 0, \ S_{t-1} = 0 \text{ and } S_{t-2} = 0 \\
2 & \text{if } S_t = 1, \ S_{t-1} = 0 \text{ and } S_{t-2} = 0 \\
3 & \text{if } S_t = 0, \ S_{t-1} = 1 \text{ and } S_{t-2} = 0 \\
4 & \text{if } S_t = 1, \ S_{t-1} = 1 \text{ and } S_{t-2} = 0 \\
5 & \text{if } S_t = 0, \ S_{t-1} = 0 \text{ and } S_{t-2} = 1 \\
6 & \text{if } S_t = 1, \ S_{t-1} = 0 \text{ and } S_{t-2} = 1 \\
7 & \text{if } S_t = 0, \ S_{t-1} = 1 \text{ and } S_{t-2} = 1 \\
8 & \text{if } S_t = 1, \ S_{t-1} = 1 \text{ and } S_{t-2} = 1.
\end{cases}
\]
Actually $L_t$ follows an 8-state Markov chain whose transition probability $\text{Prob}(S_t = j | S_{t-1} = i)$ is given by the row $j$, column $i$ element of the transition matrix

$$
P = \begin{bmatrix}
q & 0 & 0 & 0 & q & 0 & 0 & 0 \\
1 - q & 0 & 0 & 0 & 1 - q & 0 & 0 & 0 \\
0 & 1 - p & 0 & 0 & 0 & 1 - p & 0 & 0 \\
0 & p & 0 & 0 & 0 & p & 0 & 0 \\
0 & 0 & q & 0 & 0 & 0 & q & 0 \\
0 & 0 & 1 - q & 0 & 0 & 0 & 1 - q & 0 \\
0 & 0 & 0 & 1 - p & 0 & 0 & 0 & 1 - p \\
0 & 0 & 0 & p & 0 & 0 & 0 & p \\
\end{bmatrix}.
$$

In order to estimate the parameters and latent variables in Model, we can still use the Bayesian MCMC method based on the Gibbs sampler. The steps in the Gibbs sampler to generate the samples for the regular parameters and the latent stock return volatility are the same as in the Gibbs samplers for SVLR model and industrial production model. Caution should be given to generate the samples from the posterior condition distribution of $f(S_1, \cdots, S_T | H, Y^+)$, where $H$ denotes the stock volatilities all through time $T$, $Y^+$ represents all the data for the industrial production and the regular parameters are omitted for the notation convenience. Since

$$
f(S_1, \cdots, S_T | H, Y^+) = f(S_T | \overline{H}_T, \overline{Y}^+_T) \cdot \prod_{t=1}^{T-1} f(S_t | S_{t+1}, \overline{H}_t, \overline{Y}^+_t),
$$

where $\overline{H}_t = (h_1, \cdots, h_t)'$ and $\overline{Y}^+_t = (y^+_1, \cdots, y^+_t)'$, we will adopt the Forward Filtering and Backward Sampling (FFBS) method to generate the samples of $\{S_t\}$ either, i.e. doing filtering sequentially from the beginning and then sampling backward from time $T$. 

107
Specially, the Hamilton Filter for obtaining the filtered probability of latent variable $L_t$ will be iterated firstly:

\[
P(L_t = j|\overline{H}_t, \overline{Y}_t^+) = \frac{f(L_t = j|\overline{H}_{t-1}, \overline{Y}_{t-1}^+)}{\sum_{j=1}^{8} f(L_t = j|\overline{H}_{t-1}, \overline{Y}_{t-1}^+)} \cdot \Pr(h_t, y_t^+|L_t = j, \overline{H}_{t-1}, \overline{Y}_{t-1}^+)
\]

\[
P(L_{t+1} = j|\overline{H}_t, \overline{Y}_t^+) = \sum_{i=1}^{8} P_{ji} \cdot P(L_t = i|\overline{H}_t, \overline{Y}_t^+)
\]

\[i, j = 1, \ldots, 8 \text{ and } t = 1, \ldots, T,
\]

where $\Pr(h_t, y_t^+|L_t = j, \overline{H}_{t-1}, \overline{Y}_{t-1}^+)$ is calculated as the product of $f(h_t|L_t = j, \overline{H}_{t-1})$ and $f(y_t^+|L_t = j, \overline{Y}_{t-1}^+)$. The filtered probability of latent variable $S_t$ can be obtained through the following formulas,

\[
P(S_t = 0|\overline{H}_t, \overline{Y}_t^+) = P(L_t = 1|\overline{H}_t, \overline{Y}_t^+) + P(L_t = 3|\overline{H}_t, \overline{Y}_t^+) + P(L_t = 5|\overline{H}_t, \overline{Y}_t^+)
\]

\[+ P(L_t = 7|\overline{H}_t, \overline{Y}_t^+) \text{ and}
\]

\[
P(S_t = 1|\overline{H}_t, \overline{Y}_t^+) = P(L_t = 2|\overline{H}_t, \overline{Y}_t^+) + P(L_t = 4|\overline{H}_t, \overline{Y}_t^+) + P(L_t = 6|\overline{H}_t, \overline{Y}_t^+)
\]

\[+ P(L_t = 8|\overline{H}_t, \overline{Y}_t^+).
\]

Next, we could generate the sample of $S_T$ from $P(S_T|\overline{H}_T, \overline{Y}_T^+)$ and generate the samples of $S_t$ for $t = T - 1, \ldots, 1$ from $P(S_t|S_{t+1}, \overline{H}_t, \overline{Y}_t^+)$ where
8.3 Empirical Studies

In this section, we will apply Model A, Model B and Model C to the actual data in order to see if Model B and Model C, which both consider the connection between stock market volatilities and industrial productions, could improve identifying and forecasting the business cycle. The stock returns $r_t$ in these empirical studies are the monthly log difference of the S&P 500 index, which are good representatives of the overall stock market return levels. The time spell is from the January of 1985 to the December of 2011 and the data length is 324. The output growth measure $y_t^+$ in this analysis is the monthly change in the natural logarithm of the Federal Reserve Board’s index of industrial production from the January of 1985 to the December of 2011.

Firstly, let us see the case of Model A, where the process followed by stock returns is completely unrelated to industrial production. The Bayesian estimation processes for the SVLR model and the industrial production model are completely separate. Figure 8.1 shows the smoothed probability of latent variable $S_t^+$ being 1, i.e. $\text{Prob} (S_t^+ = 1 | y_1^+, y_2^+, \cdots, y_T^+)$, which tells how likely the real output growth rate is in the low-regime, i.e. the economy is in the recession phase. In the same figure, we could see the plots of rate of growth of industrial production, return of S&P 500 index and the squares of S&P index returns. All the data are quoted at monthly rates.
Figure 8.1 Empirical Results of Model A

(a) Monthly Rate of Output Growth (%) from 01/1985 to 12/2011

(b) S&P 500 Index Monthly Return (%) from 01/1985 to 12/2011
(c) Squared S&P 500 Index Monthly Return (%) from 01/1985 to 12/2011

(d) Smoothed Probability of $S_t^+ = 1$ from Model A, 01/1985 – 12/2011
Vertical lines in the panel (d) of Figure 8.1 denote the business cycle peak and trough dates as determined by NBER. The three recessions in the figure happened in the period of July 1990 – March 1991, March 2001 – November 2001 and December 2007 – June 2009. There exists correspondence between the econometric inferences from Model A and the NBER dating of economic recessions, even though the smoothed probability of $S_t^+ = 1$ associated with the first two recessions is small. The reason is that the high strength of the third recession scales down the relative strengths of the first two. The discrepancy appears around year 2005, at which the smoothed probability of $S_t^+ = 1$ is high but there is no recession taking place actually.

We could see the empirical results of Model B in Figure 8.2. In Model B, $S_t = S_t^+$, i.e. the factor that causes a shift in stock market volatility is exactly the same factor that drives the economy to go into a recession. The raw data for output and stock returns used in the analysis of Model B are also shown in the figure 8.2. The smoothed probability of $S_t$ being 1 from Model B is $\text{Prob}(S_t = 1|r_1, r_2, \ldots, r_T, y_{1T}^+, y_{2T}^+, \ldots, y_{T_T}^+)$, which tells how likely the stock market would stay in the high-volatility regime, or the industrial production fall in the low growth rate regime, since $S_t = S_t^+$. We could see that the econometric inference of business cycle from Model B is almost the same as the inference of business cycle from Model A. There still exists a false identification of business cycle from the analysis of Model B. In the following, we will study if the bivariate Model C could improve identifying and forecasting the business cycle.
Figure 8.2 Empirical Results of Model B

(a) Monthly Rate of Output Growth (%) from 01/1985 to 12/2011

(b) S&P 500 Index Monthly Return (%) from 01/1985 to 12/2011
(c) Squared S&P 500 Index Monthly Return (%) from 01/1985 to 12/2011

(d) Smoothed Probability of $S_t = 1$ from Model B, 01/1985 – 12/2011
Figure 8.3 Empirical Results of Model C

(a) Monthly Rate of Output Growth (%) from 01/1985 to 12/2011

(b) S&P 500 Index Monthly Return (%) from 01/1985 to 12/2011
(c) Squared S&P 500 Index Monthly Return (%) from 01/1985 to 12/2011

(d) Smoothed Probability of $S_t = 1$ from Model C, 01/1985 – 12/2011
Model C specifies that $S_{t-1} = S_t^+$, which means even though the forces driving the stock market and the industrial production are same, yet the stock market will be affected one period before the industrial production makes reaction. The panel (d) in Figure 8.3 plots the smoothed probability $Prob \left( S_t = 1 | r_1, r_2, \cdots, r_T, y_1^+, y_2^+, \cdots, y_T^+ \right)$, which denotes the inference based on Model C about how likely the economy was in the high stock volatility regime at period $t$, also industrial production was in the low output growth regime at period $t + 1$. The smoothed probability is based on the full sample data, including both stock index return and rate of output growth data, from January 1985 to December 2011. The correspondence between the smoothed inference and the NBER economic recession dates is remarkable. The smoothed probability line obtained from Model C is more smooth than from both Model A and Model B. The false inference about the probability of recession from both Model A and Model B around year 2005 does not appear in the analysis results any way from Model C. We actually consider the situation, in which the economic driving forces affect the stock market two periods of time before they affect industrial production, i.e. $S_{t-2} = S_t^+$, but this more complex model form could not improve the capability of identifying the business cycle. Therefore, Model C should be the optimal model form under the bivariate framework we described.
Appendix

The derivations of \( h'_t(S_t, m) \) and \( \pi(S_t, m) \) in Section 6.1.1

Since \( g^*(h_t) = \exp \left[ -\frac{1}{2} h_t - \frac{y_t^2}{2} \exp(-\hat{h}_t) (1 + \hat{h}_t - h_t) \right] \) and \( f(h_t|S_t, h_{t-1|t-1}^{(m)}) \) has a normal density, the product of \( g^*(h_t) \) and \( f(h_t|S_t, h_{t-1|t-1}^{(m)}) \cdot P(S_t|S_{t-1|t-1}^{(m)}) \) could be specified as

\[
g^*(h_t) \cdot f(h_t|S_t, h_{t-1|t-1}^{(m)}) \cdot P(S_t|S_{t-1|t-1}^{(m)}) =
\exp \left[ -\frac{1}{2} h_t - \frac{y_t^2}{2} \exp(-\hat{h}_t) (1 + \hat{h}_t - h_t) \right] \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma^2 (1 - \rho^2)} \cdot \exp \left[ -\frac{(h_t - \Xi)^2}{2\sigma^2 (1 - \rho^2)} \right] \cdot P(S_t|S_{t-1|t-1}^{(m)}),
\]

where \( \Xi = \mu_0 + \mu_1 S_t + \phi h_{t-1|t-1}^{(m)} + \rho \sigma y_{t-1} \exp(-h_{t-1|t-1}^{(m)}/2) \).

The product could be expressed as proportional to

\[
g^*(h_t) \cdot f(h_t|S_t, h_{t-1|t-1}^{(m)}) \cdot P(S_t|S_{t-1|t-1}^{(m)}) \propto
\exp \left[ -\frac{h_t}{2} + \frac{y_t^2}{2} \exp(-\hat{h}_t) h_t \right] \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma^2 (1 - \rho^2)} \cdot \exp \left[ -\frac{(h_t - \Xi)^2}{2\sigma^2 (1 - \rho^2)} \right] \cdot P(S_t|S_{t-1|t-1}^{(m)})
\]

\[
= \frac{1}{\sqrt{2\pi} \cdot \sigma^2 (1 - \rho^2)}
\cdot \exp \left[ -\frac{\sigma^2 (1 - \rho^2)[1 - y_t^2 \exp(-\hat{h}_t)] h_t + h_t^2 - 2\Xi h_t + \Xi^2}{2\sigma^2 (1 - \rho^2)} \right]
\cdot P(S_t|S_{t-1|t-1}^{(m)})
\]

118
\[
\begin{align*}
&= \frac{1}{\sqrt{2\pi \cdot \sigma^2(1-\rho^2)}} \cdot \exp \left[ - \frac{h_t^2 - 2 \left( \frac{\sigma^2}{2}(1-\rho^2) \left( 1 - y_t^2 \exp(-\hat{h}_t) \right) \right) h_t + \Xi^2}{2\sigma^2(1-\rho^2)} \right] \\
& \quad \cdot P\left( S_t \mid S_{t-1|t-1}^{(m)} \right) \\
&= \frac{1}{\sqrt{2\pi \cdot \sigma^2(1-\rho^2)}} \cdot \exp \left[ - \frac{\left( h_t - \left( \frac{\sigma^2}{2}(1-\rho^2) \left( 1 - y_t^2 \exp(-\hat{h}_t) \right) \right) \right)^2 + \Xi^2 - \left( \frac{\sigma^2}{2}(1-\rho^2) \left( 1 - y_t^2 \exp(-\hat{h}_t) \right) \right)^2}{2\sigma^2(1-\rho^2)} \right] \\
& \quad \cdot P\left( S_t \mid S_{t-1|t-1}^{(m)} \right) \\
&= \frac{1}{\sqrt{2\pi \cdot \sigma^2(1-\rho^2)}} \cdot \exp \left[ - \frac{\left( h_t - h_t^*(S_t, m) \right)^2}{2\sigma^2(1-\rho^2)} \right] \cdot \exp \left[ - \frac{\Xi^2 - \left( h_t^*(S_t, m) \right)^2}{2\sigma^2(1-\rho^2)} \right] \\
& \quad \cdot P\left( S_t \mid S_{t-1|t-1}^{(m)} \right) \\
&= f_N \left( h_t \mid h_t^*(S_t, m), \sigma^2(1-\rho^2) \right) \cdot \pi(S_t, m),
\end{align*}
\]

where

\[ h_t^*(S_t, m) = \mu_0 + \mu_1 S_t + \phi h_{t-1|t-1}^{(m)} + \rho \sigma y_{t-1} \exp(-h_{t-1|t-1}^{(m)})/2 + \frac{\sigma^2(1-\rho^2)}{2} \left[ y_t^2 \exp(-\hat{h}_t) - 1 \right], \]

\[ \pi(S_t, m) \]

\[ = \exp \left[ - \frac{\left( \mu_0 + \mu_1 S_t + \phi h_{t-1|t-1}^{(m)} + \rho \sigma y_{t-1} \exp(-h_{t-1|t-1}^{(m)})/2 \right)^2 - \left( h_t^*(S_t, m) \right)^2}{2\sigma^2(1-\rho^2)} \right] \]

\[ \cdot P\left( S_t \mid S_{t-1|t-1}^{(m)} \right). \]
References


