STRUCTURAL BREAKS AND FORECASTING IN EMPIRICAL FINANCE AND MACROECONOMICS

by

Zhongfang He

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Economics
University of Toronto

Abstract

STRUCTURAL BREAKS AND FORECASTING
IN EMPIRICAL FINANCE AND MACROECONOMICS

Zhongfang He

Doctor of Philosophy

Graduate Department of Economics

University of Toronto

2009

This thesis consists of three essays in empirical finance and macroeconomics. The first essay proposes a new structural-break vector autoregressive model for predicting real output growth by the nominal yield curve. The model allows for the possibility of both in-sample and out-of-sample breaks in parameter values and uses information in historical regimes to make inference on out-of-sample breaks. A Bayesian estimation and forecasting procedure is developed which accounts for the uncertainty of both structural breaks and model parameters. I discuss dynamic consistency when forecasting recursively and provide a solution. Applied to monthly US data, I find strong evidence of breaks in the predictive relation between the yield curve and output growth. Incorporating the possibility of structural breaks improves out-of-sample forecasts of output growth.

The second essay proposes a sequential Monte Carlo method for estimating GARCH models subject to an unknown number of structural breaks. We use particle filtering techniques that allow for fast and efficient updates of posterior quantities and forecasts in real-time. The method conveniently deals with the path dependence problem that arises in these type of models. The performance of the method is shown to work well using simulated data. Applied to daily NASDAQ returns, we find strong evidence of structural breaks in the long-run variance of returns. Models with flexible return distributions such as t-innovations or with jumps indicate fewer breaks than models with normal return

innovations and are favored by the data.

The third essay proposes a new tilt stochastic volatility model which extends the existing volatility models by modeling the asymmetric correlation between return and volatility innovations in a unified and flexible framework. The Efficient Importance Sampling (EIS) procedure is adapted to estimate the model. Simulation studies show that the Maximum Likelihood (ML)-EIS estimation of the model is accurate. The new model is applied to the CRSP daily returns. I find the extensions are significant and incorporating them improves the accuracy of volatility estimates.

Acknowledgements

I am grateful to my main advisor, Professor John Maheu, for his invaluable insights and guidance. My doctoral experience at the University of Toronto would not be the same without his exceptionally generous intellectual support and inspiration. I am also indebted to Professor Tom McCurdy, Professor Chuan Goh and Professor Christian Gourieroux for their very kind help and encouragement and agreeing to serve on my thesis committee and to be my references.

I have been privileged to benefit from the friendly and intellectually challenging community of the Department of Economics at the University of Toronto. I would like to thank Alex Maynard, Angelo Melino, Martin Burda, Victor Aguirregabiria, Michael Baker, Joanne Roberts and Michelle Alexopoulos for their helpful comments at various stages of my doctoral research.

The second essay is based on joint work with John. His guidance and input are gratefully acknowledged.

Contents

List of Tables

List of Figures

Chapter 1

Forecasting Output Growth by the Yield Curve: The Role of Structural Breaks

1.1 Introduction

Forecasting real economic activity such as output growth is an importance issue in empirical economics. Research over the last few decades has found that the nominal yield curve contains important predictive information for subsequent real economic growth. Examples of this voluminous literature include Harvey (1989), Laurent (1988,1989), Stock and Watson (1989), Chen (1991), Estrella and Hardouvelis (1991), Plosser and Rouwenhorst (1994), Davis and Fagan (1997), Estrella and Mishkin (1997,1998), Hamilton and Kim (2002) and Ang, Piazzesi, and Wei (2006), among many others.

Recently there has been growing evidence that the relationship between the yield curve and subsequent economic growth may be unstable over time; see, for example, Stock and Watson (1999,2003), Estrella, Rodrigues, and Schich (2003) and Giacomini and Rossi (2006). This forecasting instability poses a challenge for predicting output

growth by the yield curve. If the predictive relation experiences structural breaks in the past, it may change in the future as well. Ignoring the possibility of future structural breaks could result in biased and poor forecasts of output growth. This chapter is the first in the literature that studies the problem of forecasting output growth with the yield curve information in the presence of both in-sample and out-of-sample structural instability.

In this chapter, we take into account the possibility that structural breaks in the predictive relation between the yield curve and output growth have occurred in the past as well as the possibility that they may occur in the future. In contrast to the existing work on the stability of the yield curve's predictive ability for output growth using univariate models, we jointly model the dynamics of the output growth and the yield curve by a tri-variate vector autoregressive (VAR) model of output growth, the short rate and the term spread in conjunction with structural breaks. The short rate is the nominal interest rate on short maturity government debt while the term spread is the difference between nominal interest rates on long and short maturity government debts. They are commonly used in the literature to capture the predictive information in the yield curve for real economic activity. Stock and Watson (2003) provides some economic motivations for the use of these two interest rate variables for forecasting output growth. Ang et al. (2006) shows that VAR forecasts of quarterly GDP growth using the short rate and term spread are more accurate than univariate regressions of output growth on these two interest rate variables at all horizons considered in their paper. Given the well-documented evidence of structural changes in interest rates, e.g. Gray (1996), Ang and Bekaert (2002), Bansal and Zhou (2002) and Pesaran, Pettenuzzo, and Timmermann (2006), jointly modeling the structural breaks of the yield curve variables and output growth in a VAR model helps avoid attributing structural changes in interest rates to the predictive relation between the yield curve and output growth.

Built on the work of Chib (1998) and Pesaran et al. (2006), we model the structural

break process as a hierarchical hidden Markov chain[1]. The parameters of the VAR model may take different values in different break segments and are assumed to be drawn from a common meta distribution. As data in a new regime becomes available, the meta distribution is updated by Bayes rule. Hence information in the parameters of previous break segments is used to learn about parameters in the new regime in an efficient way. Forecasts are made by integrating out the uncertainty about both the in-sample and out-of-sample breaks and parameters. A Markov chain Monte Carlo (MCMC) algorithm is developed to estimate the structural break VAR model, which extends the Pesaran et al. (2006) method for univariate settings to multivariate models. We provide a careful discussion of the prior on the number of in-sample regimes implied by the hierarchical structure of this type of model. This is in contrast to Pesaran et al. (2006) which imposes a uniform prior on the number of in-sample regimes and hence is inconsistent with hierarchical priors on other model parameters. We also discuss the issue of dynamic consistency, that is, the compatibility of assumptions through time concerning the possible number of structural breaks, when forecasting recursively with structural break models in the way of Pesaran et al. (2006). A new forecasting approach is proposed that guarantees the dynamic consistency in recursive forecasting with structural break models.

The proposed model is applied to the monthly US data from January 1964 to December 2006. We consider a full-break specification in which all parameters of the VAR model are subject to structural breaks as well as a partial-break specification in which only the intercept and covariance matrix of the VAR model have structural breaks. We find that the full-break model is favored by the data despite the greater parsimony of the partial-break model. A full-break specification with 2 in-sample breaks provides the best description of the data among the models considered in this chapter.

The break dates are identified at October 1979 and January 1983, which coincides

[1]A partial sample of the alternative Bayesian models that allow for parameter shifts of random magnitude and timing includes McCulloch and Tsay (1993), Giordani and Kohn (2006), Koop and Potter (2007), Maheu and Gordon (2008) and Maheu and McCurdy (2009), among others.

closely with the change in monetary policy regime with the advent of the US Fed chairman Volcker in late 1979. Before 1979, the short rate predicts output growth while the term spread is largely insignificant. The regime between 1979 and 1983 is marked by exceptionally high volatilities of all three variables, during which neither the short rate nor the term spread is able to predict output growth. The most recent regime since 1983 has much lower volatility. During this period, the predictive power of the short rate largely disappears while the term spread becomes significant. This new finding is in contrast to the studies that do not consider the possibility of structural breaks. Most of the studies in the literature find that the short rate has little marginal predictive content for output growth once spreads are included, e.g. Plosser and Rouwenhorst (1994) and Stock and Watson (2003). In contrast, Ang et al. (2006) finds that the short rate has more predictive power for quarterly GDP growth than term spreads. This chapter, by taking into account the possibility of parameter shifts, finds that the relative importance of the short rate and term spread is changing over time and the spread has more predictive power for output growth than the short rate in the most recent regime.

We perform recursive out-of-sample density forecasting exercises from January 2002 to December 2006 to compare the performance of the proposed structural break VAR model with the conventional no-break VAR model. The results show that incorporating the possibility of structural breaks significantly improves the forecasting accuracy of output growth from 1 to 12 months ahead. The improvements in forecasting accuracy are steady and almost continuous throughout the forecasting period. We also experiment with imposing informative priors on the structural break VAR model for forecasting since a number of studies (Litterman (1980,1986),Kadiyala and Karlsson (1997)) have advocated their use in Bayesian VAR forecasting. We find that in our forecasting exercise, informative priors do not necessarily lead to better forecasts than the benchmark priors.

The rest of the chapter is organized as follows. Section 1.2 describes the structural break VAR model and the estimation method. Section 1.3 explains the forecast proce-

dure. The empirical estimates are presented in Section 1.4. Section 1.5 is the conclusions. Technical details of the estimation algorithm are presented in the appendices.

1.2 The Model

We consider a tri-variate VAR model of the output growth, the short rate and the term spread for forecasting output growth, which has been found to produce superior out-of-sample forecasts of GDP growth than univariate OLS regressions (Ang et al. (2006)).

The VAR model is assumed to be subject to a random number of structural breaks which separate different regime segments. A subset of the VAR coefficients may change their values in different regimes. Formally, the structural break VAR model is

$$y_t = \mu_{s_t} + \Phi_{s_t} y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \Sigma_{s_t}) \tag{1.1}$$

where $y_t \equiv (g_t, r_t, x_t)'$, g_t is the output growth, r_t is the short rate and x_t is the term spread at time t, $t = 1, 2, ..., T$.

The regimes are indexed by a state variable $s_t \in \{1, 2, \ldots, K\}$ following the transition matrix

$$\pi = \begin{pmatrix} \pi_{1,1} & 1-\pi_{1,1} & 0 & ... & 0 \\ 0 & \pi_{2,2} & 1-\pi_{2,2} & ... & 0 \\ & & & ... & \\ 0 & 0 & ... & \pi_{K-1,K-1} & 1-\pi_{K-1,K-1} \\ 0 & 0 & ... & 0 & 1 \end{pmatrix} \tag{1.2}$$

where $\pi_{i,j}$ is the probability of moving to regime j given that the current regime is i. Note that at each point of time, the state variable s_t can either stay in the current regime or jump to the next one. The transition terminates in regime K. We will denote $\pi_i \equiv \pi_{i,i}$ for notational simplicity. A break occurs at time t if $s_t \neq s_{t-1}$. Note $s_t = K$ implies that K regimes, or, $K-1$ breaks, have occurred in the data up to time t.

This formulation of structural breaks builds on the original work of Chib (1998), which has been used in many applications, e.g. Pastor and Stambaugh (2001), Kim, Morley, and Nelson (2005), Pesaran et al. (2006), Liu and Maheu (2008) and He and Maheu (2008). The regime-switching model of Hamilton (1988) can be viewed as a special case of this setup if identical states are assumed to recur (Pesaran et al. (2006)).

Chib (1998) has developed an efficient Bayesian MCMC algorithm for estimating this class of structural break models. But the Chib method can not handle the possibility of out-of-sample breaks during forecasting horizons and hence is not well suited for out-of-sample forecasts unless one is willing to assume that no new breaks could occur out of sample.

To perform forecasts while taking into account possible out-of-sample breaks, we need to model the underlying process of the parameters in different regimes. In this chapter, we follow Pesaran et al. (2006) and posit hierarchical priors for the regime parameters. Parameters in each regime are assumed to be drawn from a common meta distribution. As data from new regimes become available, the meta distribution is updated by Bayes rule. Hence information in the parameters of previous regimes is used to learn about parameters in the new regime in an efficient way. This method has the attractive feature that it retains the sampling efficiency of the Chib method while conveniently modeling the underlying process of the regime parameters and hence is well suited for out-of-sample forecasts. We develop a new MCMC algorithm to estimate this class of models, extending the Pesaran et al. (2006) method for univariate settings to multivariate models.

Let $\phi_k = vec\left([\mu_k \;\; \Phi_k]'\right)$ be the vector containing the elements of μ_k and Φ_k. We assume that the linear coefficients ϕ_k, the covariance matrices Σ_k and the transition probabilities π_k are independently drawn from the following distributions respectively.

$$\phi_k \sim N(b_0, B_0)$$

$$\Sigma_k \sim IW(\Omega_0, v_0 + 3)$$

for $k = 1, 2, ..., K$, where IW denotes the inverse Wishart distribution, and

$$\pi_k \sim Beta(\alpha_0, \beta_0)$$

for $k = 1, 2, ..., K-1$.

At the next level of the hierarchy, we assume that

$$b_0 \sim N(a_0, A_0)$$

$$B_0 \sim IW(D_0, d_0)$$

$$\Omega_0 \sim IW(\Psi_0, f_0)$$

$$v_0 \sim Gamma(\rho_0, \lambda_0)$$

$$\alpha_0 \sim Gamma(q_0, \gamma_0)$$

$$\beta_0 \sim Gamma(r_0, \delta_0)$$

where a_0, A_0, D_0, d_0, Ψ_0, f_0, ρ_0, λ_0, q_0, γ_0, r_0 and δ_0 are hyper-parameters and are specified *a priori*. This hierarchical structure creates dependence between parameters in different regimes. Given an estimation sample, model parameters from different in-sample regimes ϕ_k, Σ_k and π_k are used to update the distributions of the hierarchical prior parameters b_0, B_0, Ω_0, v_0, α_0 and β_0 by Bayes rule. Inference on parameters of possible out-of-sample regimes is then based on the updated distributions of these hierarchical prior parameters. Hence information contained in in-sample estimates is efficiently used to produce forecasts outside the estimation sample.

To conduct Bayesian estimation, we divide the parameters into 3 blocks for a given number of in-sample regimes K: the latent states $S = (s_1, s_2, ..., s_T)$, parameters of the hierarchical priors $\Theta_0 = (b_0, B_0, \Omega_0, v_0, \alpha_0, \beta_0)$ and the other model parameters $\Theta = (\phi_1, ..., \phi_K, \Sigma_1, ..., \Sigma_K, \pi_1, ..., \pi_{K-1})$. A Gibbs sampler is developed to estimate this hierarchical structural break model, which iterates sampling from the following conditional distributions

- $S|\Theta_0, \Theta$
- $\Theta_0|S, \Theta$
- $\Theta|S, \Theta_0$

The details of the algorithm are provided at Appendix A.1.

Inference on the number of in-sample regimes K is conducted based on the posterior distribution $p(s_T = K|Y_T),\ K = 1, 2, ..., \overline{K}$, since according to the transition matrix of Equation (1.2), the number of in-sample regimes K equals the state variable at the end of data sample s_T. In theory, the possible number of in-sample regimes $\overline{K}$ can be as large as the number of observations T, i.e. a structural break at every period of time. But in practice, it can often be set to be a relatively small number $\overline{K} < T$ provided that the posterior distribution $p(s_T = K|Y_T)$ does not support going beyond $\overline{K}$ in-sample regimes.

Applying Bayes rule, the posterior distribution $p(s_T = K|Y_T)$ can be decomposed as

$$p(s_T = K|Y_T) \propto p(s_T = K)p(Y_T|s_T = K)$$

where $p(s_T = K)$ is the prior probability and $p(Y_T|s_T = K)$ is the marginal likelihood of Y_T given K in-sample regimes. There is a significant Bayesian literature on methods of computing the marginal likelihoods, e.g. Gelfand and Dey (1994), Newton and Raftery (1994), Chib (1995), Fruhwirth-Schnatter (1995, 2004), Meng and Wong (1996), and Chib and Jeliazkov (2001). Miazhynskaia and Dorffner (2006) provides a nice comparison of the various methods of computing marginal likelihoods. In this chapter, we adopt the modified harmonic mean method of Gelfand and Dey (1994) which has been found to be accurate (Miazhynskaia and Dorffner (2006)) while computationally convenient. The details of implementing the modified harmonic mean method for the structural break VAR model are provided in Appendix A.2.

The prior $p(s_T = K)$ on the number of in-sample regimes K is implied by the hierarchical prior on transition probabilities π_k, $k = 1, 2, ..., \overline{K} - 1$, as

$$p(s_T = K) = \int p(s_T = K|\pi_1, ..., \pi_{\overline{K}-1}) \prod_{k=1}^{\overline{K}-1} p(\pi_k|\alpha_0, \beta_0) p(\alpha_0, \beta_0) d\pi_1 \cdots d\alpha_0 d\beta_0$$

This distribution has no closed form but can be computed by using the approximation

$$p(s_T = K) \approx \frac{1}{n} \sum_{i=1}^{n} I\{s_T^{(i)} = K\}$$

where $s_T^{(i)}$ is a draw of s_T from the prior $p(s_T)$ and the indicator function $I\{s_T^{(i)} = K\} = 1$ if $s_T^{(i)} = K$ and 0 otherwise. To draw from the prior $p(s_T)$, one can sample a path of states $\{1, s_2^{(i)}, s_3^{(i)}, \cdots, s_T^{(i)}; s_T^{(i)} \leq \overline{K}\}$ conditional on each draw of $\alpha_0^{(i)} \sim Gamma\,(q_0, \gamma_0)$, $\beta_0^{(i)} \sim Gamma\,(r_0, \delta_0)$, $\pi_k^{(i)} \sim Beta\left(\alpha_0^{(i)}, \beta_0^{(i)}\right)$, $i = 1, 2, ..., n$, and then keep $s_T^{(i)}$ as a sample from $p(s_T)$. This practice is in contrast to Pesaran et al. (2006) which imposes $p(s_T = K) = 1/\overline{K}$ and hence is inconsistent with the actual prior $p(s_T = K)$ implied by the hierarchical priors on other model parameters.

1.3 Out-of-Sample Forecasts

In this section, we show how to produce out-of-sample forecasts of output growth from the model proposed in Section 1.2. Given the posterior draws of parameters and latent states based on time T information $Y_T \equiv \{y_1, y_2, ..., y_T\}$, we forecast future values of output growth, which is the first element of the vector y, by taking into account the uncertainty about both in-sample and out-of-sample breaks and parameters.

Let h be the forecasting horizon. The predictive distribution of output growth g_{T+h} integrates out uncertainty about the number of in-sample breaks by Bayesian model averaging

$$p(g_{T+h}|Y_T) = \sum_{K=1}^{\overline{K}} p(g_{T+h}|Y_T, s_T = K) p(s_T = K|Y_T) \tag{1.3}$$

where K is the number of in-sample regimes in the data Y_T and has the upper limit $\overline{K}$. The weight used in the averaging, $p(s_T = K|Y_T)$, is the posterior probability

of the structural break VAR model with K in-sample regimes. The other ingredient $p(g_{T+h}|Y_T, s_T = K)$ is the predictive distribution of output growth conditional on K in-sample regimes.

To integrate out the uncertainty about out-of-sample breaks, the conventional method, e.g. Pesaran et al. (2006), applies the decomposition

$$p(g_{T+h}|Y_T, s_T = K) = \sum_{j=K}^{K+h} p(g_{T+h}|Y_T, s_T = K, s_{T+h} = j)p(s_{T+h} = j|Y_T, s_T = K) \quad (1.4)$$

that is, conditional on K in-sample regimes at time T, it is assumed that up to $K + h$ regimes could occur at time $T + h$. This approach is reasonable when viewed statically. But once being put in a recursive forecasting context as in practice, it becomes logically inconsistent if the upper limit of in-sample regimes $\overline{K}$ is kept fixed throughout. To see this, consider the example $K = \overline{K}$ and $h = 1$. When making forecasts at time T, one assumes that $s_{T+1} = \overline{K} + 1$ is possible according to Equation (1.4). But after arriving at $T + 1$, one assumes $s_{T+1} \leq \overline{K}$ as the maximum number of in-sample regimes in the data Y_{T+1} is fixed at $\overline{K}$, which is inconsistent with the assumptions made at time T. This creates a dynamic inconsistency problem: assumptions concerning the possible number of structural breaks are not consistent through time when forecasting recursively. Increasing the upper limit $\overline{K}$ by 1 when moving 1 period forward is not an attractive solution since the extra computation cost would soon become too high to be practical and the number of possible structural breaks entertained should not be unboundedly increasing.

In this chapter, we propose to set an upper limit on the total number of *both in-sample and out-of-sample regimes*

$$p(g_{T+h}|Y_T, s_T = K) = \sum_{j=K}^{\min\{\overline{K}, K+h\}} p(g_{T+h}|Y_T, s_T = K, s_{T+h} = j)p(s_{T+h} = j|Y_T, s_T = K) \quad (1.5)$$

in order to be dynamically consistent. When forecasting h-periods ahead at any time T, the new method guarantees that the number of regimes assumed for time $T + h$ satisfies $s_{T+h} \leq \overline{K}$ regardless of the current regime s_T. This will be consistent with the

assumption after one actually arrives at time $T+h$ that the upper limit of the number of in-sample regimes s_{T+h} is $\overline{K}$. This simple modification solves the dynamic inconsistency problem suffered by the conventional method while entertaining no extra computation cost.

When viewed as a mixture distribution, sampling from the predictive distribution of output growth $p(g_{T+h}|Y_T)$ in Equation (1.3) is straightforward:

Step 1. Compute the posterior probabilities $p(s_T = K|Y_T)$ for $K = 1, 2, ..., \overline{K}$.

Step 2. Sample indices $k \in \{1, 2, ..., \overline{K}\}$ from a multinomial distribution with the posterior probabilities $p(s_T = K|Y_T)$ as parameters.

Step 3. If $k = K$, then sample $g_{T+h}^{(i)}$ from the predictive distribution $p(g_{T+h}|Y_T, s_T = K)$ according to Equation (1.5).

The resulting sample of output growth $\{g_{T+h}^{(i)}\}_{i=1}^{n}$ will provide a complete distribution of the future output growth based on the current information set. The predictive mean of any function of the output growth $f(g_{T+h})$ can be consistently estimated as

$$E[f(g_{T+h})|Y_T] \approx \frac{1}{n}\sum_{i=1}^{n} f(g_{T+h}^{(i)})$$

There are two important ingredients for this forecasting method. One is the posterior probability of in-sample regimes $p(s_T = K|Y_T)$, whose computation has been discussed in detail in the preceding section of model description. Note that since $p(s_T = K|Y_T) = \int p(s_T = K, s_{T-1}, ..., s_2|Y_T)ds_{T-1} \cdots ds_2$, this probability integrates over the paths of the states, i.e. the locations of in-sample breaks.

The other ingredient is the predictive distribution given $s_T = K$,

$$p(g_{T+h}|Y_T, s_T = K) = \sum_{j=K}^{\min\{\overline{K}, K+h\}} p(g_{T+h}|Y_T, s_T = K, s_{T+h} = j)p(s_{T+h} = j|Y_T, s_T = K)$$

which is used in Step 3 above. For ease of exposition, consider first the predictive distribution conditional on the relevant model parameters. Let θ denote the set containing the

in-sample parameters ϕ_K, Σ_K and out-of-sample parameters π_K, π_{K+1}, ϕ_{K+1}, Σ_{K+1},..., $\pi_{\min\{\overline{K},K+h\}}$, $\phi_{\min\{\overline{K},K+h\}}$, $\Sigma_{\min\{\overline{K},K+h\}}$, that is,

$$\theta = \left(\phi_K, \Sigma_K, \pi_K, \pi_{K+1}, \phi_{K+1}, \Sigma_{K+1}, ..., \pi_{\min\{\overline{K},K+h\}}, \phi_{\min\{\overline{K},K+h\}}, \Sigma_{\min\{\overline{K},K+h\}}\right)$$

We have

$$\begin{aligned} & p(g_{T+h}|Y_T, s_T = K, \theta) \\ = & \sum_{j=K}^{\min\{\overline{K},K+h\}} p(g_{T+h}|Y_T, s_T = K, s_{T+h} = j, \theta) p(s_{T+h} = j|Y_T, s_T = K, \theta) \end{aligned} \tag{1.6}$$

Consider the component distribution

$$p(g_{T+h}|Y_T, s_T = K, s_{T+h} = j, \theta) p(s_{T+h} = j|Y_T, s_T = K, \theta), \quad j \geq K$$

of Equation (1.6), which specifies a total of j regimes at time $T+h$ with K in-sample regimes and $j-K$ out-of-sample regimes. For the case $j > K$, i.e. $j-K$ out-of-sample breaks occurring during $T+1, ..., T+h$, it is necessary to integrate over all possible locations of the out-of-sample breaks. Let τ_k be the location of the k-th break point. Applying the law of total probability, we have

$$\begin{aligned} & p(g_{T+h}|Y_T, s_T = K, s_{T+h} = j, \theta) p(s_{T+h} = j|Y_T, s_T = K, \theta) \\ = & \sum_{1 \leq i_1 < i_2 ... < i_{j-K} \leq h} p(\tau_K = T + i_1, \tau_{K+1} = T + i_2, ..., \tau_{j-1} = T + i_{j-K}|Y_T, s_T = K, \theta) \cdot \\ & p(g_{T+h}|Y_T, s_T = K, \tau_K = T + i_1, \tau_{K+1} = T + i_2, ..., \tau_{j-1} = T + i_{j-K}, \theta) \end{aligned} \tag{1.7}$$

This integrates over all possible locations of out-of-sample breaks. The probability of the out-of-sample break scenario in the first term of Equation (1.7) is given by

$$\begin{aligned} & p(\tau_K = T + i_1, \tau_{K+1} = T + i_2, ..., \tau_{j-1} = T + i_{j-K}|Y_T, s_T = K, \theta) \\ = & \pi_K^{i_1 - 1}(1 - \pi_K)\pi_{K+1}^{i_2 - i_1 - 1}(1 - \pi_{K+1}) \cdots \pi_j^{h - i_{j-K}} \end{aligned} \tag{1.8}$$

The clearest way of writing out the predictive densities under each scenario of out-of-sample breaks

$$p(g_{T+h}|Y_T, s_T = K, \tau_K = T + i_1, \tau_{K+1} = T + i_2, ..., \tau_{j-1} = T + i_{j-K}, \theta)$$

which is the second term in Equation (1.7), is to use recursive equations. Since the densities under specific scenarios of out-of-sample breaks are Gaussian, we only need to specify their means and variances. Let $\mu(i)$, $\Sigma(i)$ be the mean and covariance matrix of the predictive distribution of y_{T+i} from the structural break VAR model of Equation (1.1). We have

$$\begin{aligned}
\mu(1) &= \mu_{s_{T+1}} + \Phi_{s_{T+1}} y_T, \\
\Sigma(1) &= \Sigma_{s_{T+1}}, \\
\mu(i) &= \mu_{s_{T+i}} + \Phi_{s_{T+i}} \mu(i-1), \\
\Sigma(i) &= \Sigma_{s_{T+i}} + \Phi_{s_{T+i}} \Sigma(i-1) \Phi'_{s_{T+i}}
\end{aligned} \tag{1.9}$$

where s_{T+i} is the index of the regime at $T+i$, $i = 1, 2, ..., h$. Let $e \equiv (1 \;\; 0 \;\; 0)'$. It can be shown (Lutkepohl (2006)) that the marginal density of output growth g_{T+h}, which is the first element of the vector y_{T+h}, is

$$g_{T+h}|Y_T, s_T = K, \tau_K, \tau_{K+1}, ..., \tau_{j-1}, \theta \sim N(e'\mu(h), e'\Sigma(h)e) \tag{1.10}$$

where $\mu(h)$ and $\Sigma(h)$ track the specific scenario of out-of-sample breaks τ_K,τ_{K+1},...,τ_{j-1} by tracking the out-of-sample states $s_{T+1}, ..., s_{T+h}$ according to Equation (1.9). For the case $j = K$, i.e. no out-of-sample breaks, it is straightforward to show

$$p(s_{T+h} = K|Y_T, s_T = K, \theta) = \pi_K^h \tag{1.11}$$

and the predictive density $p(g_{T+h}|Y_T, s_T = K, s_{T+h} = K, \theta)$ can be obtained from Equations (1.9) and (1.10) by setting $s_{T+1} = \cdots = s_{T+h} = K$.

When there is uncertainty surrounding the parameters, the predictive distribution $p(g_{T+h}|Y_T, s_T = K)$ needs to integrate over the parameters.

$$\begin{aligned}
& p(g_{T+h}|Y_T, s_T = K) \\
=& \int p(g_{T+h}|Y_T, s_T = K, \theta) p(\theta|Y_T, s_T = K) d\theta \\
\approx& \frac{1}{n} \sum_{i=1}^{n} p(g_{T+h}|Y_T, s_T = K, \theta^{(i)})
\end{aligned} \tag{1.12}$$

where based on posterior draws $\{\alpha_0^{(i)}, \beta_0^{(i)}, b_0^{(i)}, B_0^{(i)}, \Omega_0^{(i)}, v_0^{(i)}, \phi_K^{(i)}, \Sigma_K^{(i)}\}_{i=1}^n$ from the structural break VAR model given data Y_T and K in-sample regimes, the components in $\theta^{(i)}$ are $\pi_k^{(i)} \sim Beta(\alpha_0^{(i)}, \beta_0^{(i)})$ for $k = K, ..., \min\{\overline{K}, K+h\}$, $\phi_k^{(i)} \sim N(b_0^{(i)}, B_0^{(i)})$ and $\Sigma_k^{(i)} \sim IW(\Omega_0^{(i)}, v_0^{(i)})$ for $k = K+1, ..., \min\{\overline{K}, K+h\}$. Computing the conditional predictive density $p(g_{T+h}|Y_T, s_T = K, \theta^{(i)})$ of Equation (1.12) follows discussions in the preceding paragraph. To sample from this distribution, one can first simulate a path of out-of-sample regimes $\{s_{T+1}^{(i)},...,s_{T+h}^{(i)}\}$ based on the transition probabilities $\pi_K^{(i)},...,\pi_{\min\{\overline{K},K+h\}}^{(i)}$ and then sample $g_{T+h}^{(i)}$ from the predictive distribution under the simulated path of regimes according to Equations (1.9) and (1.10).

1.4 Empirical Results

We use monthly nominal zero-coupon yield data with maturities of 3 months and 5 years from the Center for Research in Security Prices (CRSP) spanning January 1964 to December 2006. Prior to this period, there were few traded long bonds. So data on long yields before 1964 may be unreliable (Fama and Bliss (1987)). The short rate is the 3-month yield from the Fama-Bliss risk-free rate file. The term spread is constructed as the 5-year yield minus the 3-month yield, with the 5-year yield data derived from the Fama-Bliss discount bond file. Ang et al. (2006) finds that this term spread has the best predictive power for quarterly GDP growth among all the maturities of spreads considered in their paper. All yield data are continuously compounded. Data on real output growth is the log growth rate of industrial production index from the FRED dataset of the US Federal Reserve. Table 1.1 provides summary statistics of the sample. A plot of the data is presented in Figure 1.1. All data are scaled up by 100 in estimation.

1.4.1 In-Sample Estimates

First, we conduct an in-sample study of the full-sample data from January 1964 through December 2006. The purpose is to identify if there are structural breaks in the predictive content of the yield curve for output growth and, if so, how the predictive relations change over time.

Let $m = 3$ be the number of variables in the structural break VAR model. The priors are set to be: $b_0 \sim N(0, 100I_{m(m+1)})$, $B_0 \sim IW(I_{m(m+1)}, m(m+1)+4)$, $\Omega_0 \sim IW(0.001I_{m(m+1)}, m+4)$, $v_0 \sim Gamma(2,3)$, $\alpha_0 \sim Gamma(20,1)$ and $\beta_0 \sim Gamma(2, 0.05)$, where $I_{m(m+1)}$ denotes a $m(m+1) \times m(m+1)$ identity matrix. These priors are diffuse over realistic ranges of values for the parameters. We set the upper limit on the number of in-sample regimes $\overline{K}$ to be 5. As will be seen below, the posterior distribution on the number of in-sample regimes does not support going beyond 5. Table 1.2 contains the simulated prior distribution on the number of in-sample regimes implied by the hierarchical prior of transition probabilities. It can be seen that the prior probability $p(s_T = 1) = 0.733$ and hence the prior strongly favors no break in the data. We discard 5,000 initial draws and retain the next 80,000 for posterior analysis. The chain mixes well. The acceptance rates of the Metropolis-Hastings steps are all in the range between 0.3 and 0.5. As an example, the posterior draws of the predictive coefficient of the term spread in the first regime is presented in Figure 1.2.[2]

We estimate the no-break VAR model as well as the full-break VAR model of Equation (1.1) with 1 to 4 in-sample breaks, that is, $K = 1, 2, ..., 5$. The marginal likelihoods peak at 2 in-sample breaks and diminishes as more breaks are introduced. The resulting log marginal likelihoods are presented in Table 1.3. It can be seen that the full-break VAR model with 2 in-sample breaks has the largest marginal likelihood. Table 1.2 provides

[2]For the favored model of 3 in-sample regimes, the numerical standard errors (NSE) (Geweke (2005)) of all 155 model parameters are less than 0.6% of the estimated posterior means. 82% of the NSEs are less than 0.3% of the estimated posterior means.

the posterior distribution of the number of in-sample regimes K along with its simulated prior distribution. The posterior probability $p(s_T = 3|Y_T)$ is 0.998. So there is strong evidence supporting 3 in-sample regimes or 2 in-sample breaks. The large difference in the prior and posterior distributions of K suggests that the data is informative about the number of in-sample regimes. Table 1.4, 1.5 and 1.6 provide the parameter estimates of the full-break VAR model with 3 in-sample regimes. The slope coefficients of the full-break VAR model exhibit smaller differences across regimes than the intercepts and covariance matrices. So a partial-break VAR model in which only the intercept and the covariance matrix have structural breaks is also investigated. We find that the log marginal likelihoods of the partial break VAR model with 1, 2 and 3 in-sample breaks are -1044.06, -1004.06 and -1000.45 respectively, significantly below those of the corresponding full-break VAR models.

Based on the estimates of marginal likelihoods, the full-break VAR model with 2 in-sample breaks is found to provide the best description of the data among the models considered. The posterior median of the first break date is October 1979 with the 95% credible set between July 1979 and October 1979 while the posterior median of the second break date is January 1983 with the 95% credible set between November 1982 and March 1985. The posterior distributions of the break dates are plotted in Figure 1.3. These estimates of break dates are broadly consistent with the findings of previous studies using univariate models. For example, Estrella et al. (2003) suggests a break around September 1983 when forecasting monthly industrial production growth by term spread. Pesaran et al. (2006) finds evidence of breaks in an AR(1) model of monthly 3-month T-bill rates at September 1979 and September 1982. Maheu and Gordon (2008) finds evidence of break in an AR(2) model of quarterly GDP growth at the 3rd quarter of 1983.

There seems to be a compelling connection of the break dates to the change in the monetary policy regime with the advent of the US Fed chairman Volcker in the late 1979. Some fundamental changes in the Fed's operating procedure took place beginning

at October of that year. The move to a monetary policy regime targeting money growth was essentially over by 1983. These dates coincide closely with the estimates of break dates. As will be seen below, the behavior and relation of the data series vary greatly across regimes identified by these break dates.

Estimates of the parameters for the full-break VAR model with 3 in-sample regimes are presented in Table 1.4, 1.5 and 1.6. Parameters significant at the 95% level, that is, the value of 0 is outside the 95% credible set, are marked by "*" for visual convenience. For comparison, estimates of the corresponding no-break VAR model are presented in Table 1.7. It can be seen that, for the structural break VAR model, the covariances between the innovations of the variables change over the 3 regimes and the second regime is marked by high volatility of all three variables. Before this regime, it is the short rate that significantly predicts output growth while the term spread is insignificant. During the high-volatility regime, none of these variables can predict the output growth. After the high-volatility regime, the predictive power of the short rate disappears while the term spread becomes significant. Figure 1.4 plots the posterior distributions of the predictive coefficients of the short rate and term spread in the 3 in-sample regimes. It can be seen that the posterior distributions of these coefficients have undergone noticeable changes during the sample period. In contrast, the no-break VAR estimates show that both the short rate and term spread are significant over the whole sample period. Note that most of the studies of predicting output growth involving short rates and term spreads found that the short rate has little marginal predictive power once spreads are included, e.g. Plosser and Rouwenhorst (1994) and Stock and Watson (2003). In contrast, Ang et al. (2006) found that the short rate has more predictive power for GDP growth than term spreads. Our approach differs from the existing studies by explicitly modeling the possible structural instability of the predictive relation and finds that the relative importance of short rates and term spreads is changing over time. In the most recent regime, the term spread has more predictive power for output growth than the short rate.

1.4.2 Performance of Out-of-Sample Forecasts

To assess the usefulness of incorporating structural breaks in the predictive relation of the yield curve and output growth, a recursive out-of-sample forecasting exercise is conducted to compare the performance of the structural break VAR model with the no-break VAR model, that is, at each point of time, only historically available information is used to estimate the models and make forecasts. We focus on the forecast of output growth as it is the major interest of this chapter. Forecasts of all 3 variables from the VAR models can be done in a similar way. We use the same diffuse priors as in the in-sample study of Section 1.4.1. The estimates are based on 80,000 posterior draws after discarding 5,000 initial draws.

We consider 4 forecasting horizons: 1-month-ahead, 3-months-ahead, 6-months-ahead and 12-months-ahead, which are commonly used in practice. In forecasting, we assume that there are at most 2 out-of-sample breaks during these forecasting horizons. This assumption is plausible since an out-of-sample regime shorter than the considered forecasting horizons seems unlikely based on historical estimates. As a practical matter, scenarios of more than 2 out-of-sample breaks have numerically negligible probabilities as estimates of the transition probabilities π_k are uniformly close to 1.

The log predictive Bayes factor for output growth is used to assess the models' forecasting performances. Given a sample of data y_1, y_2,...,y_T and the starting forecast date τ, $\tau \leq T-h$, the log cumulative predictive likelihood of output growth is

$$\sum_{t=\tau}^{T-h} \log\left(p(g_{t+h}|Y_t)\right) \tag{1.13}$$

where h is the forecast horizon. Note that the starting forecast date $\tau \leq T-h$ since $T-h$ is the last point in which a h-period ahead forecast can be evaluated given T data points. For two competing models M_1 and M_2, the difference in their log cumulative predictive likelihoods of output growth

$$\sum_{t=\tau}^{T-h} \log\left(\frac{p(g_{t+h}|Y_t; M_1)}{p(g_{t+h}|Y_t; M_2)}\right) \tag{1.14}$$

is the log predictive Bayes factor of output growth. Model M_1 is favored by the data if the log predictive Bayes factor is positive. This measure keeps the cumulative record of out-of-sample forecasting performance of the models and is the central criterion in Bayesian model comparison (Geweke and Whiteman (2005)). As can be seen from Section 1.3, the predictive distribution of output growth $p(g_{t+h}|Y_t)$ under the structural break VAR model is a mixture of normal distributions with different means and variances and hence is likely to be highly non-Gaussian. Compared with traditional measures such as the Root Mean Squared Error (RMSE), the predictive Bayes factor can provide a more complete view of the forecasting performance in such cases.

The model averaging approach of Equation (1.3) is used to compute the predictive likelihoods of output growth $p(g_{t+h}|Y_t)$ according to the formulas in Section 1.3, which integrate out the uncertainty about both in-sample and out-of-sample breaks and parameters. For each forecasting horizon h, the component models of Equation (1.3) are the full-break VAR models with $K = 1$ to 5 in-sample regimes. For each component model with K in-sample regimes, we compute the posterior probability $p(s_t = K|Y_t)$ and the predictive density $p(g_{t+h}|Y_t, s_t = K)$ which integrates over all possible scenarios of out-of-sample breaks according to Equations (1.6) and (1.12). The sum of the product of these predictive densities $p(g_{t+h}|Y_t, s_t = K)$ and posterior probabilities $p(s_t = K|Y_t)$ is computed as in Equation (1.3) and is recorded as the predictive likelihood of output growth $p(g_{t+h}|Y_t)$. This procedure is repeated recursively from January 2002 to December 2006 for a 5-year period. Similarly we compute the predictive likelihoods of output growth $p(g_{t+h}|Y_t)$ recursively for the no-break VAR model. The log predictive Bayes factor of output growth is then computed based on the predictive likelihoods according to Equations (1.13) and (1.14).

Figure 1.5 plots the log predictive Bayes factors of the full-break VAR model against the no-break VAR model. It can be seen that the full-break VAR model outperforms the no-break VAR model in all of the 4 forecasting horizons, despite the fact that the no-break

model is more parsimonious. The predictive Bayes factors increase steadily throughout the forecasting period[3]. At the end of the 5-year forecasting period, the log predictive Bayes factors are more than 4.5 for all of the horizons, which translate into cumulative predictive likelihoods of output growth from the structural break VAR model more than 90 times higher than those from the no-break VAR model. This forecasting exercise illustrates the precision gain of moving from a no-break VAR model to the structural break VAR representation.

We experiment with imposing informative priors on the structural break VAR model since in the Bayesian literature of VAR forecasting, the use of informative priors has been frequently advocated, e.g. Kadiyala and Karlsson (1997). One popular choice is the Litterman prior (Litterman (1980,1986)). Kadiyala et al. (1997) specifies the Litterman prior as a Gaussian prior distribution for parameters in the intercept and slope matrix of VAR models. The mean of the prior is 1 for diagonal elements of the slope matrix and equals 0 for the intercept and off-diagonal elements of the slope matrix. The variances of the prior recommended in Kadiyala et al. (1997) are set as: $10^5\sigma_i^2$ for the i-th element in the intercept, 0.05 for the diagonal elements and $\frac{0.005\sigma_i^2}{\sigma_j^2}$ for the off-diagonal (i,j) element of the slope matrix, where σ_i is the residue standard error of an autoregression for variable i. The covariances in the prior are set to be 0.

In the context of the structural break VAR model, the relevant hyper-parameters in the hierarchical priors are a_0, A_0, d_0 and D_0 as defined in Section 1.2. We set a_0 to be the mean of the Litterman prior[4], $A_0 = 0.1I_{m(m+1)}$, $d_0 = m(m+1)+50$ and D_0 the covariance matrix of the Litterman prior scaled by $d_0 - m(m+1) - 1$, where $m = 3$ is the number

[3]There is a drop in the predictive Bayes factors at the forecast of output growth in Sep 2005. This can be explained by a sharp change of output growth around this period. The output growth rate is -1.64% in August and jumps to 1.14% in September 2005, which are the lowest and highest output growth rates during the forecasting period from Jan 2002 to Dec 2006. The average output growth rate during the forecasting period is only 0.21%. Nevertheless forecasts of the break VAR model pick up momentum shortly after this period.

[4]For the diagonal element in the slope matrix of VAR corresponding to output growth, the mean is set to be 0.3 instead of 1 since, unlike the short rate and term spread which are highly persistent, the 1st order autocorrelation of output growth is about 0.3.

of variables in the VAR. These values of hyper-parameters calibrate the distribution for parameters in the intercepts and slope matrices of the structural break VAR model to center around the Litterman priors with small variation. Table 1.8 reports the log cumulative predictive likelihood from Jan 2002 to Dec 2006 for the structural break VAR model with this informative prior along with those from the structural break VAR model with diffuse prior and no-break VAR model. Among the structural break VAR models, the informative prior produces better 1-month-ahead forecasts while being outperformed in 6-months-ahead forecasts by the diffuse prior. For the horizons 3 and 12 months, performances of the two priors are close. Nevertheless both produce superior forecasts than the no-break VAR model.

1.5 Conclusion

In this chapter, we present a new structural break VAR model for predicting real output growth by the nominal yield curve which allows for the possibility of both in-sample and out-of-sample breaks in parameter values. We jointly model structural breaks in the output growth and yield curve. A Bayesian estimation approach is provided which extends the method of Pesaran et al. (2006) for univariate models to multivariate settings. We provide a discussion of the prior on the number of in-sample regimes implied by the hierarchical structure of this type of model. This is in contrast to Pesaran et al. (2006) which imposes a uniform prior on the number of in-sample regimes and hence is inconsistent with hierarchical priors on other model parameters. A new forecasting method is proposed which guarantees dynamic consistency when forecasting output growth recursively in real time. The empirical application focuses on the monthly US data from 1964 to 2006. We find strong evidence of structural breaks in the predictive relation between the yield curve and output growth in late 1979 and early 1983. Before 1979, the short rate has more predictive power for output growth than the term spread while the term spread

becomes more significant since the break of 1983. In the forecasting exercise, we find that incorporating the possibility of structural breaks produces more accurate out-of-sample forecasts of output growth than the no-break VAR model at all horizons considered in this chapter.

Table 1.1: Summary Statistics of Data

	Mean	Std. Deviation	Skewness	Kurtosis
Output Growth	0.253	0.728	-0.668	6.455
Short Rate	5.901	2.788	1.025	4.722
Term Spread	1.011	1.138	-0.470	3.312

This table reports the summary statistics of the monthly US data from January 1964 through December 2006.

Table 1.2: Distributions of the Number of In-Sample Regimes

	$K = 1$	$K = 2$	$K = 3$	$K = 4$	$K = 5$
prior $p(s_T = K)$	0.733	0.181	0.054	0.019	0.013
posterior $p(s_T = K\|Y_T)$	0	0	0.998	0.002	0

This table reports the prior and posterior distributions of the number of in-sample regimes in the full-break VAR model for the monthly US data from January 1964 through December 2006.

Table 1.3: Model Comparison by Marginal Likelihoods

	No-Break	1-Break	2-Break	3-Break	4-Break
log ML	-1131.81	-1008.69	-978.27	-983.65	-992.68

This table reports the log marginal likelihoods of the no-break VAR model and the full-break VAR models with 1 to 4 in-sample breaks for the monthly US data from January 1964 through December 2006.

Table 1.4: Parameter Estimates of the Full-Break VAR Model: Regime 1

$$\mu_1 = \begin{pmatrix} 0.567^* \\ (0.106, 1.040) \\ 0.039 \\ (-0.259, 0.335) \\ 0.114 \\ (-0.127, 0.355) \end{pmatrix}$$

$$\Phi_1 = \begin{pmatrix} 0.321^* & -0.072^* & 0.084 \\ (0.188, 0.454) & (-0.145, -0.001) & (-0.050, 0.218) \\ 0.035 & 0.993^* & 0.042 \\ (-0.044, 0.116) & (0.947, 1.038) & (-0.041, 0.124) \\ -0.004 & -0.012 & 0.917^* \\ (-0.069, 0.060) & (-0.049, 0.025) & (0.850, 0.983) \end{pmatrix}$$

$$\Sigma_1 = \begin{pmatrix} 0.621^* & & \\ (0.505, 0.762) & & \\ -0.013 & 0.218^* & \\ (-0.068, 0.040) & (0.178, 0.268) & \\ -0.005 & -0.134^* & 0.142^* \\ (-0.049, 0.038) & (-0.169, -0.105) & (0.115, 0.174) \end{pmatrix}$$

This table reports the posterior means and 95% credible sets for regime 1 parameters of the full-break VAR model with 2 in-sample breaks $y_t = \mu_{s_t} + \Phi_{s_t} y_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0, \Sigma_{s_t})$, where y_t=(output growth, short rate, term spread), for the monthly US data from January 1964 through December 2006. Parameters significant at the 95% level are marked by "*".

Table 1.5: Parameter Estimates of the Full-Break VAR Model: Regime 2

$$\mu_2 = \begin{pmatrix} 0.414 \\ (-0.364, 1.233) \\ 0.196 \\ (-0.511, 0.959) \\ 0.153 \\ (-0.510, 0.823) \end{pmatrix}$$

$$\Phi_2 = \begin{pmatrix} 0.351^* & -0.046 & 0.041 \\ (0.064, 0.632) & (-0.115, 0.021) & (-0.131, 0.210) \\ 0.449^* & 0.981^* & 0.093 \\ (0.061, 0.852) & (0.911, 1.049) & (-0.141, 0.337) \\ -0.295^* & -0.008 & 0.832^* \\ (-0.577, -0.025) & (-0.068, 0.049) & (0.663, 0.994) \end{pmatrix}$$

$$\Sigma_2 = \begin{pmatrix} 0.860^* & & \\ (0.526, 1.387) & & \\ 0.508^* & 1.949^* & \\ (0.102, 1.091) & (1.121, 3.196) & \\ -0.268 & -1.082^* & 0.865^* \\ (-0.632, 0.003) & (-1.830, -0.604) & (0.526, 1.407) \end{pmatrix}$$

This table reports the posterior means and 95% credible sets for regime 2 parameters of the full-break VAR model with 2 in-sample breaks $y_t = \mu_{s_t} + \Phi_{s_t} y_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0, \Sigma_{s_t})$, where y_t=(output growth, short rate, term spread), for the monthly US data from January 1964 through December 2006. Parameters significant at the 95% level are marked by "*".

Table 1.6: Parameter Estimates of the Full-Break VAR Model: Regime 3

$$\mu_3 = \begin{pmatrix} 0.112 \\ (-0.086, 0.331) \\ -0.001 \\ (-0.010, 0.097) \\ 0.091 \\ (-0.023, 0.214) \end{pmatrix}$$

$$\Phi_3 = \begin{pmatrix} 0.115 & 0.003 & 0.075^* \\ (-0.005, 0.233) & (-0.030, 0.032) & (0.006, 0.144) \\ 0.077^* & 0.990^* & 0.016 \\ (0.018, 0.135) & (0.976, 1.004) & (-0.018, 0.050) \\ 0.056 & -0.007 & 0.942^* \\ (-0.011, 0.123) & (-0.026, 0.010) & (0.903, 0.981) \end{pmatrix}$$

$$\Sigma_3 = \begin{pmatrix} 0.282^* & & \\ (0.238, 0.334) & & \\ 0.032^* & 0.071^* & \\ (0.015, 0.050) & (0.056, 0.085) & \\ 0.001 & -0.028^* & 0.094^* \\ (-0.019, 0.020) & (-0.039, -0.017) & (0.079, 0.112) \end{pmatrix}$$

This table reports the posterior means and 95% credible sets for regime 3 parameters of the full-break VAR model with 2 in-sample breaks $y_t = \mu_{s_t} + \Phi_{s_t} y_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0, \Sigma_{s_t})$, where y_t=(output growth, short rate, term spread), for the monthly US data from January 1964 through December 2006. Parameters significant at the 95% level are marked by "*".

Table 1.7: Parameter Estimates of the No-Break VAR Model

$$\mu = \begin{pmatrix} 0.307^{*} \\ (0.115, 0.495) \\ 0.021 \\ (-0.127, 0.167) \\ 0.068 \\ (-0.049, 0.185) \end{pmatrix}$$

$$\Phi = \begin{pmatrix} 0.294^{*} & -0.032^{*} & 0.058^{*} \\ (0.212, 0.376) & (-0.055, -0.008) & (0.001, 0.117) \\ 0.138^{*} & 0.989^{*} & 0.013 \\ (0.074, 0.204) & (0.970, 1.008) & (-0.033, 0.059) \\ -0.042 & 0.001 & 0.935^{*} \\ (-0.094, 0.010) & (-0.014, 0.016) & (0.898, 0.971) \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} 0.458^{*} & & \\ (0.405, 0.515) & & \\ 0.055^{*} & 0.288^{*} & \\ (0.024, 0.087) & (0.255, 0.326) & \\ -0.027^{*} & -0.160^{*} & 0.180^{*} \\ (-0.052, -0.002) & (-0.185, -0.137) & (0.160, 0.204) \end{pmatrix}$$

This table reports the posterior means and 95% credible sets for parameters of the no-break VAR model $y_t = \mu + \Phi y_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0, \Sigma)$, where y_t=(output growth, short rate, term spread), for the monthly US data from January 1964 through December 2006. Parameters significant at the 95% level are marked by "*".

Table 1.8: Comparing Log Cumulative Predictive Likelihoods of Output Growth

Forecasting Horizon	No-Break VAR	Full-Break VAR with Diffuse Prior	Full-Break VAR with Informative Prior
1 month	-55.78	-48.86	-43.72
3 months	-54.83	-46.64	-46.96
6 months	-52.50	-46.42	-47.90
12 months	-45.49	-40.81	-40.57

This table reports log cumulative predictive likelihoods of output growth $\sum_{t=T-60}^{T-h} log(p(g_{t+h}|Y_t))$, $h = 1, 3, 6, 12$, for the no-break VAR model and the full-break VAR models with diffuse and informative priors during the period Jan 2002-Dec 2006.

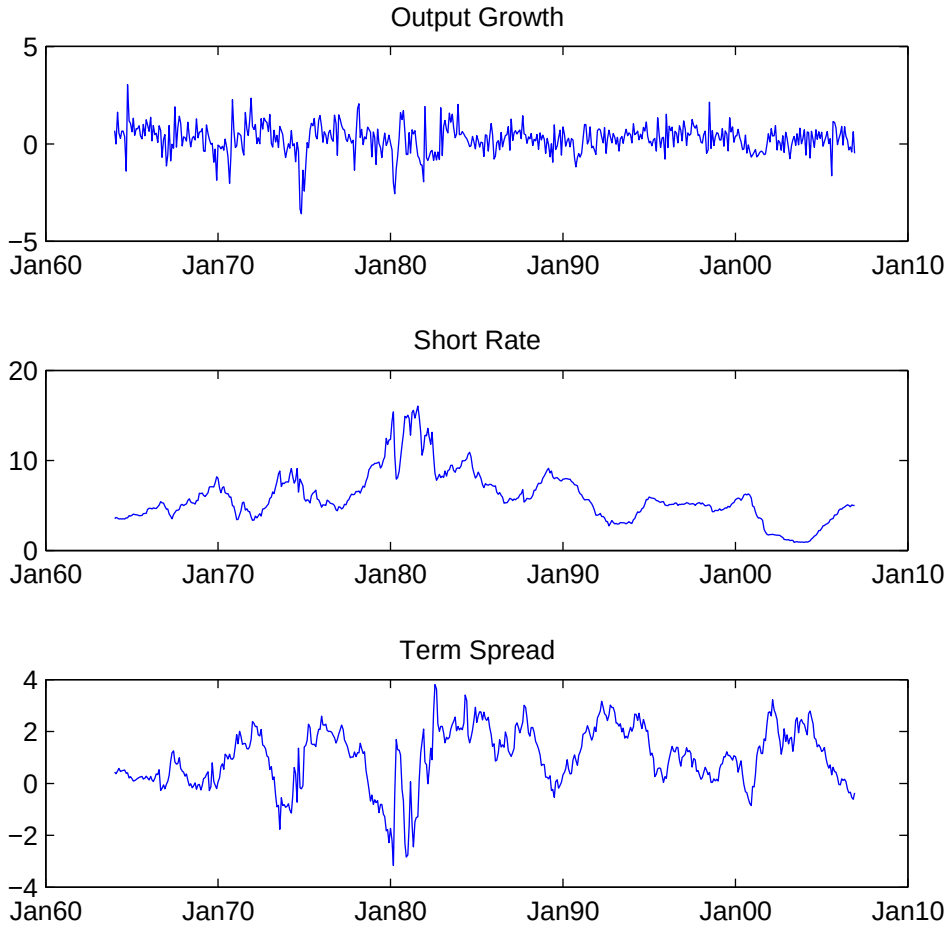

Figure 1.1: The Monthly US Data: January 1964 to December 2006

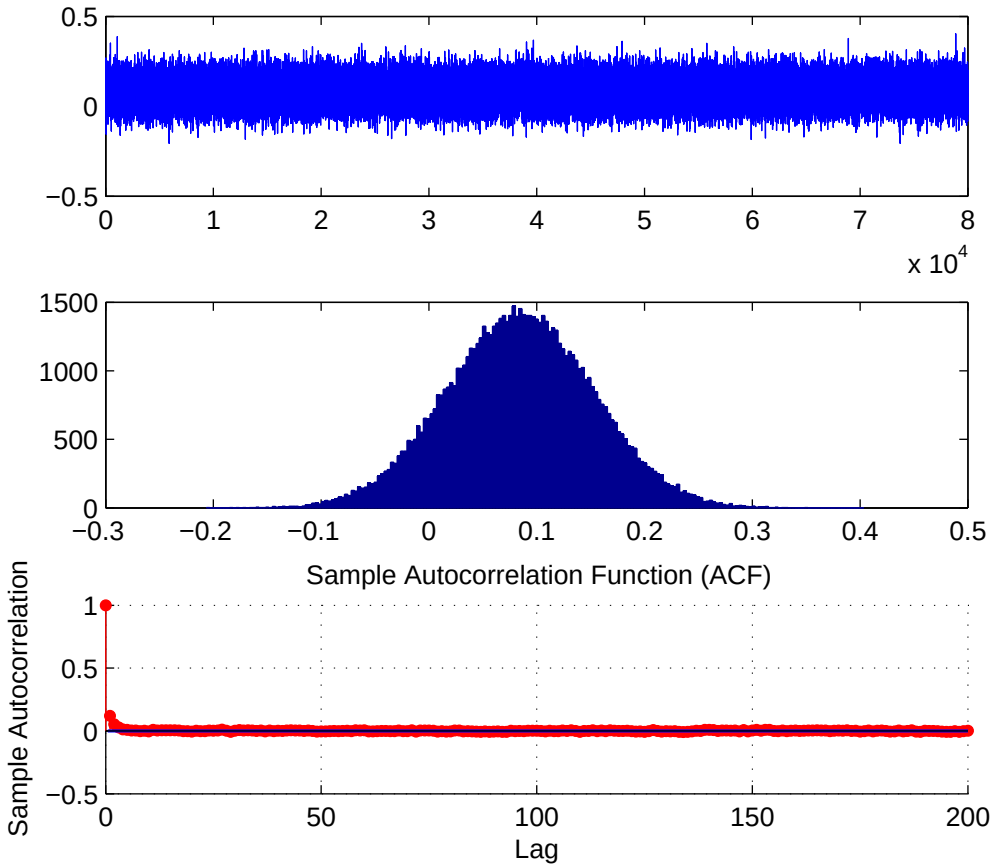

Figure 1.2: Posterior Draws of the Predictive Coefficient of Term Spread in Regime 1

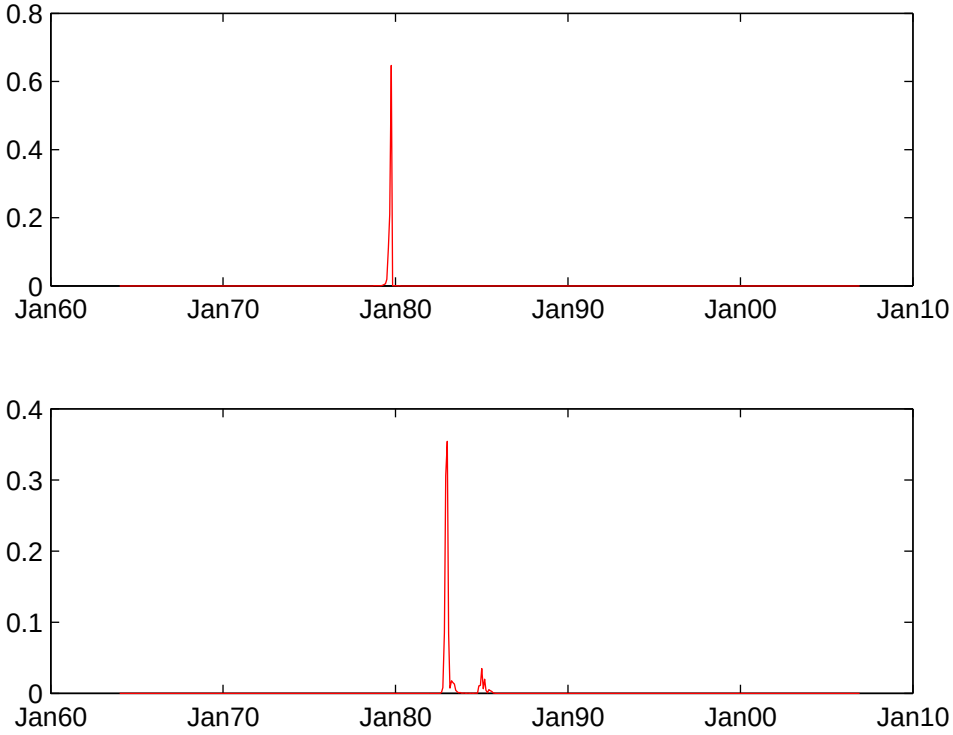

Figure 1.3: Posterior Distributions of the Break Dates

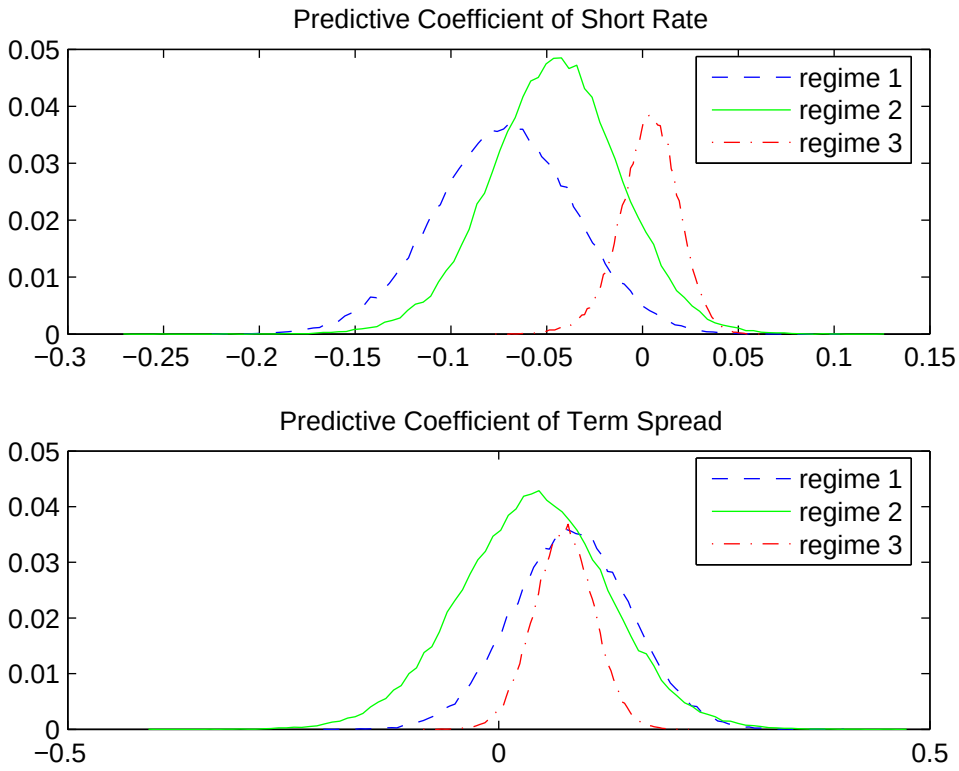

Figure 1.4: Posterior Distributions of the Predictive Coefficients of Short Rate and Term Spread

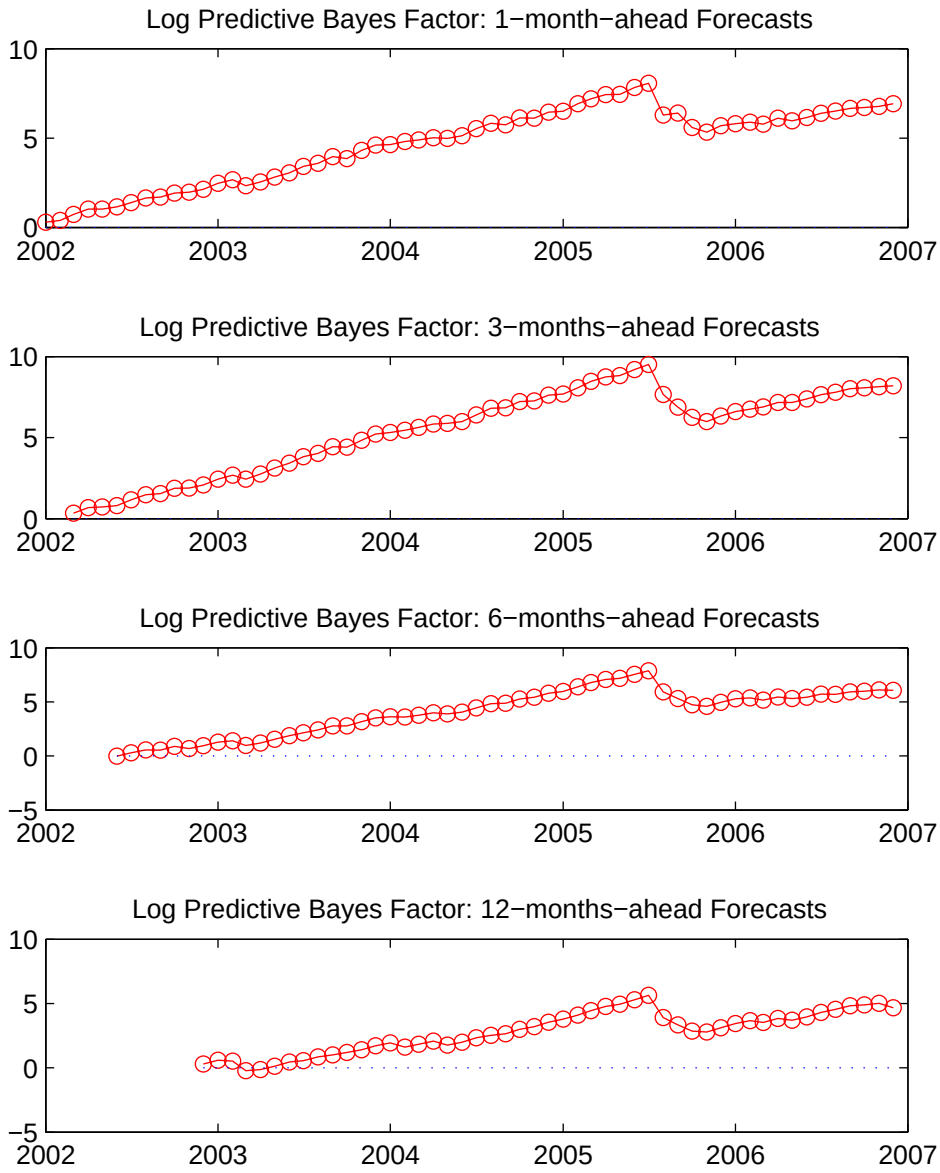

Figure 1.5: Log Predictive Bayes Factors: Break vs. No Break (Note: A positive value of log predictive Bayes factor favors the full-break VAR model.)

Chapter 2

Real Time Detection of Structural Breaks in GARCH Models

2.1 Introduction

There is growing interest among practitioners as well as academics in estimating models with structural breaks that are able to capture abrupt changes in the behavior of economic agents in response to changes in economic conditions and policies as well as political events. One prominent example of parameter change over time is the GARCH model, e.g. Lamoureux and Lastrapes (1990), Chu (1995), Mikosch and Starica (2004), Hillebrand (2005), Rapach and Strauss (2006) and Chopin (2007).[1] A common finding of this literature is that neglected parameter changes cause substantial biases in parameter estimates and poor forecasts.

This chapter addresses the econometric challenges of estimating GARCH models subject to structural breaks by using a Bayesian sequential Monte Carlo approach. The Bayesian approach delivers exact finite sample results that are well suited to dealing

[1]Other papers that find structural change in volatility dynamics include Liu and Maheu (2008) and Starica and Granger (2005).

with structural breaks that may occur after relatively short intervals. We present a particle filtering approach to sequential estimation that builds on the change-point model of Chib (1998). The approach allows the number of breaks as well as model parameters to be estimated jointly in one run and can handle models with path dependence conveniently. Our empirical application underscores the importance of model assumptions when investigating breaks. A model with normal return innovations result in strong evidence of breaks; while more flexible return distributions such as t-innovations or adding jumps to the model still favor breaks but indicate much more uncertainty regarding the time and impact of them.

Chib (1998) proposes a convenient state-space formulation of structural break models in which a discrete latent state variable drives structural breaks and follows a constrained Markov chain. An efficient MCMC algorithm for estimation and computing Bayes factors for the purpose of model selection is provided in Chib (1996,1998). One limitation of this method is that the number of possible structural breaks needs to be specified *a priori*. The model is repeatedly estimated conditional on $k = 0, 1, ..., k_{\max}$ breaks occurring. Inference on the number of break points requires computation of the marginal likelihoods for all specifications.[2]

There is a significant Bayesian literature on analysis of structural breaks using Markov chain Monte Carlo (MCMC) methods, e.g. Barry and Hartigan (1993), McCulloch and Tsay (1993), Green (1995) and Gerlach, Carter, and Kohn (2000). Recent work by Koop and Potter (2007), Maheu and Gordon (2008) and Pesaran, Pettenuzzo, and Timmermann (2006) extends this to forecasting out-of-sample in the presence of structural breaks.

[2]In the MCMC context setting $k_{\max}$ to a sufficiently large number and doing only one estimation run is not an attractive alternative. This forces there to be $k_{\max}$ breaks in-sample, which may be larger than the true number of breaks, and as a result will increase parameter uncertainty and decrease the quality of predictions. On the other hand, not enforcing that all states are visited in sampling may cause state collapsing in which no data are sorted into some states. This leads to poor convergence of the chain (Scott (2002)).

These MCMC approaches are based on the entire sample of data. Whenever a new observation arrives, the whole posterior sampling algorithm has to be re-run in order to update estimates of parameters and state variables. Computation becomes cumbersome and practically infeasible when new observations arrive in high frequency such as in many financial applications. The method proposed in this chapter allows for fast efficient updates to posterior quantities as new data arrives.

Another challenging problem for MCMC analysis of structural breaks is how to handle models with path dependence, e.g. GARCH models. For example, a change in a parameter of the conditional variance at time t will affect all future conditional variances Gray (1996). Due to this path dependence the dimension of the state space grows over time. Computing the likelihood of the model requires integrating over all possible paths of break points, which quickly become computationally intractable. As a result, the forward-backward filter of Chib (1998) which allows for an efficient MCMC approach is not available for this class of models.[3]

This chapter proposes a particle filter algorithm that can provide fast approximate updates to posterior quantities as new data arrives and estimate the number of break points simultaneously with other parameters and state variables in a single run. It can estimate models with path dependence conveniently and compute the marginal likelihoods as a by-product of estimation. By focusing on the sequential filtering problem rather than the smoothing problem (MCMC), the path dependence that structural breaks induce in GARCH models is eliminated.

A particle filter is a class of sequential Monte Carlo filtering methods that approximate the posterior distribution of the latent state variables by a set of *particles* and associated probability weights. The original particle filtering algorithm is proposed by

[3]Bauwens, Preminger, and Rombouts (2006) develops a single-move Gibbs sampler which can deal with path dependence in a Markov switching GARCH model with a fixed number of regimes. To determine the number of regimes would require computing the marginal likelihood, for which as far as we know, there is no efficient method available.

Gordon, Salmond, and Smith (1993). Important subsequent contributions include Kitagawa (1996), Liu and Chen (1998), Pitt and Shephard (1999), Carpenter, Clifford, and Fearnhead (1999), Johannes and Polson (2006) and Polson, Stroud, and Muller (2008) among others.

Chopin (2007) has developed a particle filtering algorithm for estimating structural break models in which the fixed model parameters are formulated as part of the state variables. Our approach differs in that we use Chib's formulation of a structural break model and the fixed parameters are treated separately from state variables. Parameter learning is realized through the mixture kernel smoothing method of Liu and West (2001).[4] Our formulation of structural breaks has the advantage that it can easily incorporate the case where only a subset of parameters have breaks while in the Chopin approach, it is not straightforward, if not impossible, due to the path dependence problem that occurs in local MCMC moves.[5] In contrast, we are able to compare GARCH specifications that allow breaks in all parameters with versions that allow breaks only to the intercept of the conditional variance. It is this latter specification that isolates breaks in the long-run variance and has empirical support (Starica and Granger (2005)).

We assume structural breaks are governed by a finite dimension nonhomogenous Markov chain. In contrast to the MCMC method of Chib (1998), our approach does not enforce that all states be visited. Following Chopin and Pelgrin (2004), we specify an upper bound on the number of structural breaks to facilitate computation. As long as the upper bound on the number of breaks is large enough we can jointly estimate the model parameters and the number of structural breaks at each point in time based on a single run of the algorithm.

The proposed algorithm is applied to structural break GARCH (SB-GARCH) models.

[4]The mixture kernel smoothing approach of Liu and West (2001) has been successfully used in several papers including Carvalho and Lopes (2007), Casarin and Trecroci (2006) and Raggi and Bordignon (2008).

[5]Chopin (2007) uses local MCMC sampling based on Gilks and Berzuini (2001) to rejuvenate the particles and avoid degeneracy.

We investigate 4 specifications, namely, a partial SB-GARCH model in which only the intercept of the volatility equation has breaks with normal and t return innovations, and a full SB-GARCH models in which all parameters are subject to breaks with normal and t return innovations. Based on simulated data, we find that the algorithm performs well in estimating the number and locations of breaks as well as fixed parameters. Furthermore, the marginal likelihoods, computed as a byproduct of the particle filter, can be used to identify the true model specification correctly. For empirical application, we analyze the daily NASDAQ returns from January 3, 1995 to December 29, 2006. Our benchmark GARCH model with normal innovations identifies two breaks in late 1997 and early 2004, which are associated with changes in the long-run variance of returns. Based on marginal likelihoods, a partial SB-GARCH model with t-innovations has the highest cumulative predictive power compared with full SB-GARCH models as well as no-break GARCH models. The structural break GARCH model with student-t return innovations outperforms the normal specification and indicates much more uncertainty regarding the time and impact of breaks on the model.

We find it important to use a flexible model for daily returns; failure to do so may result in the false identification of structural change. According to the filter estimates of the states the evidence of structural change with normal return innovations is largely removed once we use t-innovations or introduce jumps into the model. Nevertheless, Bayes factors favor the structural break model even with the large amount of uncertainty about breaks.

In summary, we propose a particle filtering technique that allows for fast and efficient updates of posterior quantities and forecasts of GARCH with structural breaks. The method can handle the path dependence problem in the conditional variance and can estimate specifications in which all parameters change from a break as well as the case where only a subset of parameters change. The performance of the method is analyzed using both simulated and NASDAQ return data.

The chapter is organized as follows. Section 2.2 presents the general structural break model we consider. Section 2.3 provides a brief review of particle filtering methods and the algorithm for the general structural break model. In Section 2.4, we detail the SB-GARCH models we estimate. Section 2.5 gives the simulation results of the algorithm and Section 2.6 is the empirical application on daily NASDAQ returns. The conclusions are presented in Section 2.7.

2.2 A General Structural Break Model

We consider a general structure break model in which the observation y_t is drawn from an analytically tractable density $p(y_t|D_{t-1}, \theta_t)$, where D_{t-1} is the set of information known at time $t-1$ and θ_t is the set of parameters. The changing parameters θ_t are assumed to be driven by an unobservable state variable s_t such that

$$\theta_t = \theta_k \quad \text{when } s_t = k, \;\; k \in \{1, 2, ..., \overline{K}\}$$

where θ_k is the parameter value in state k. Note some of the elements of θ_t may be constant across states. The state variable s_t is modeled as a Markov chain with the transition probability matrix

$$P = \begin{pmatrix} p_{11} & 1-p_{11} & 0 & ... & 0 \\ 0 & p_{22} & 1-p_{22} & ... & 0 \\ & & & ... & \\ 0 & 0 & ... & p_{\overline{K}-1,\overline{K}-1} & 1-p_{\overline{K}-1,\overline{K}-1} \\ 0 & 0 & ... & 0 & 1 \end{pmatrix} \tag{2.1}$$

A structural break at time t occurs when $s_t \neq s_{t-1}$. This formulation of structural breaks was originally proposed by Chib (1998) and is used extensively in subsequent papers in the literature, e.g. Pastor and Stambaugh (2001), Kim, Morley, and Nelson (2005) and Liu and Maheu (2008). It has two major benefits. First, it automatically enforces an

ordering of break points in the series y_t. Moreover, when viewed as a hidden Markov model (HMM), it facilitates the marriage with the existing large literature on HMM's and hence the development of efficient estimation methods (Scott (2002)).

Chib (1998) has developed a general MCMC algorithm for structural break models with a fixed number of states. But there are important limitations of this approach. As a smoothing algorithm, estimates of state variables rely on the information of the entire sample. It is not computationally feasible to update estimates as each new observation arrives in high frequency since the whole algorithm has to be re-run to incorporate this new information. It is inconvenient to determine the number of states via the existing MCMC methods. The usual practice is to run the algorithm repeatedly conditional on a fixed number $\overline{K}$ of states specified *a prior* and calculate their marginal likelihoods for several $\overline{K}$. The number of states is then determined by comparing these marginal likelihoods. Computing the marginal likelihood is unfortunately often a complicated issue in the MCMC context. The whole estimation process can be time-consuming and become impractical in real applications where inference needs to be updated frequently. In the case of models with path dependence such as the structural break GARCH model, which we discuss next, the Chib algorithm is not directly applicable.

2.2.1 Structural Breaks and GARCH

In this section we review some of the computational issues in estimating structural breaks in GARCH specifications and the resulting path dependence. The GARCH model of Engle (1982) and Bollerslev (1986) has been widely used in practice for estimating the volatility of asset returns. One typical form is

$$y_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim i.i.d.(0,1) \tag{2.2}$$

$$\sigma_t^2 = c + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{2.3}$$

where y_t is the return of some financial asset. The sum $\alpha+\beta$ measures the persistence of the volatility process. A common finding in the literature is that estimates of this sum tend to be close to one, indicating that the volatility is highly persistent. However, it has been argued that this high persistence may be due to structural breaks in the volatility process, e.g. see Diebold (1986), Lamoureux and Lastrapes (1990) and Mikosch and Starica (2004), and the omission of possible shifts in parameters would bias upward the estimate of persistence parameters and impair volatility forecast, see Hamilton and Susmel (1994), Gray (1996) and Klaassen (2002), among others.

Mikosch and Starica (2004) showed theoretically that structural breaks in the *unconditional variance* of the GARCH volatility process could cause spuriously high persistence estimates. This motivates specifying a partial structural break GARCH (SB-GARCH) model

$$y_t = \sigma_t \epsilon_t \tag{2.4}$$

$$\sigma_t^2 = c_{s_t} + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{2.5}$$

where s_t is the unobserved state variable driving the breaks in the volatility process. Within regime s_t, the implied long run variance is $c_{s_t}/(1-\alpha-\beta)$.

Estimating SB-GARCH models is a challenging problem since the likelihood of y_t depends on the entire sequence of past states up to time t due to the recursive structure of its volatility. For example, consider the evolution of the conditional variance given a start-up value of σ_1^2 and $s_1 = 1$. At $t = 2$ there can be a break $s_2 = 2$ in which $\sigma_2^2 = c_2 + \alpha y_1^2 + \beta \sigma_1^2$ or no break $s_2 = 1$, $\sigma_2^2 = c_1 + \alpha y_1^2 + \beta \sigma_1^2$. It is clear that the variance is a function of s_2 so we denote it as $\sigma_2^2(s_2)$. Now at time $t = 3$ there is again the possibility of a break or no break. In this case the conditional variance is a function of s_2 and s_3, i.e. $\sigma_3^2(s_2, s_3)$. In general the likelihood at time t is a function of the entire history of states $\{s_2, s_3, ..., s_t\}$. Efficient MCMC methods assume a first-order Markov chain for the discrete state (Chib (1998)) to implement the forward-backward smoother to draw the states. Therefore, $\{s_2, s_3, ..., s_t\}$ must be recast to a first-order Markov chain

whose dimension is increasing in t making the Chib (1998) approach infeasible[6]. The path dependence in the conditional variance also occurs when α and/or β change from a break.

To circumvent the path dependence problem in the context of Markov switching GARCH, a variety of alternative tractable approximations have been proposed in the literature, e.g. Gray (1996), Klaassen (2002) and Haas, Mittnik, and Paolella (2004). Bauwens, Preminger, and Rombouts (2006) develops a single-move MCMC algorithm that could be adapted to estimate SB-GARCH models with a *known fixed* number of states. However, this would not provide real time estimates, nor is it feasible to estimate SB-GARCH models with an unknown number of states via existing MCMC methods.[7]

By focusing on the sequential filtering problem rather than the smoothing problem (MCMC), the path dependence that structural breaks induce in GARCH models is eliminated.

2.3 Particle Filter

In this chapter, we propose a sequential Monte Carlo filter algorithm that can provide online real time inference and estimate the number of states simultaneously with other parameters and state variables in a single run. Out-of-sample forecasts integrate out uncertainty from both parameters and break points. The algorithm can estimate structural break models with path dependence conveniently and compute the marginal likelihood through a predictive likelihood decomposition.

[6]The particle filter approach of Chopin and Pelgrin (2004) which is based on forward backward simulation smoother is also not applicable due to the path dependence problem.

[7]As far as we know, no methods for computing marginal likelihoods of this class of models are available. So it is effectively infeasible to estimate this class of models via existing MCMC methods unless one is willing to assume that the number of break points is known a priori.

2.3.1 Particle Filter with Known Parameters

To facilitate the discussion of the estimation algorithm, we first provide a brief review of particle filters with known parameters since the particle filter was originally proposed to estimate time-varying state variables while assuming the fixed parameters are known. For detailed discussions of particle filter, see the books edited by Doucet, de Freitas, and Gordon (2001) and Ristic, Arulampalam, and Gordon (2004).

The foundational particle filtering algorithm is proposed by Gordon, Salmond, and Smith (1993). Roughly speaking, the particle filter is a class of sequential Monte Carlo filtering methods which approximate the posterior distribution of the state variables, $p(s_t|D_t)$, by a set of *particles* $\{s_t^{(i)}\}_{i=1}^N$, with probability weights $\{w_t^{(i)}\}_{i=1}^N$, where $\sum_{i=1}^N w_t^{(i)} = 1$. This relation is conventionally denoted as

$$\{s_t^{(i)}, w_t^{(i)}\}_{i=1}^N \sim p(s_t|D_t).$$

Given a set of particles and weights, the posterior mean of any function of the state variable $f(s_t)$ can be directly estimated as

$$E[f(s_t)|D_t] \approx \sum_{i=1}^N f(s_t^{(i)}) w_t^{(i)}$$

The predictive density is approximated as

$$p(s_{t+1}|D_t) = \int p(s_{t+1}|s_t)p(s_t|D_t)ds_t \approx \sum_{i=1}^N p(s_{t+1}|s_t^{(i)})w_t^{(i)} \tag{2.6}$$

and the filtering density is approximated as

$$p(s_{t+1}|D_{t+1}) \propto p(y_{t+1}|s_{t+1})p(s_{t+1}|D_t) \approx p(y_{t+1}|s_{t+1}) \sum_{i=1}^N p(s_{t+1}|s_t^{(i)})w_t^{(i)}. \tag{2.7}$$

The centerpiece of a particle filter algorithm is how to propagate particles forward from $\{s_t^{(i)}, w_t^{(i)}\}_{i=1}^N$ to $\{s_{t+1}^{(i)}, w_{t+1}^{(i)}\}_{i=1}^N$. Gordon, Salmond, and Smith (1993) employ a sampling/importance resampling (SIR) scheme. Though intuitive and straightforward to implement, the SIR filter is vulnerable to the weight degeneracy problem, that is, only

a small subset of the particles is assigned appreciable weights in the propagation stage and hence the effective size of particles is reduced, which leads to greater approximation errors.

Many extensions on the basic SIR filter have been proposed in the literature, e.g. see Kitagawa (1996), Carpenter, Clifford, and Fearnhead (1999) and Johannes and Polson (2006). In this chapter, we focus on the auxiliary particle filter (APT) developed in Pitt and Shephard (1999) which is widely used and reliable.

The basic idea of the APT is to treat Equation (2.7) as a mixture distribution and introduces an auxiliary variable r to index the mixture components

$$p(s_{t+1}, r|D_{t+1}) \propto p(y_{t+1}|s_{t+1})p(s_{t+1}|s_t^{(r)})w_t^{(r)}$$

Given a set of particles and weights $\{s_t^{(i)}, w_t^{(i)}\}_{i=1}^N$, the sequential scheme first samples the indices $\{r^i\}_{i=1}^N$, $r^i \in \{1, 2, ..., N\}$, by using the weights $p(y_{t+1}|\widehat{s}_{t+1}^{(i)})w_t^{(i)}$ where $\widehat{s}_{t+1}^{(i)}$ is a predictive value of $s_{t+1}^{(i)}$ such as the mode or the mean of $p(s_{t+1}|s_t^{(i)})$. Then $s_{t+1}^{(i)}$ is sampled from $p(s_{t+1}|s_t^{r^i})$ and the new weights are computed as $w_{t+1}^{(i)} \propto p(y_{t+1}|s_{t+1}^{(i)})/p(y_{t+1}|\widehat{s}_{t+1}^{r^i})$. By giving more importance to particles with large predictive values, the APF improves sampling efficiency while significantly reducing weight degeneracy problems.

2.3.2 Particle Filter with Unknown Parameters

In general, the model parameters are unknown and have to be estimated from the observed data. There have been a number of useful approaches of learning model parameters θ proposed in the literature. Following an idea in Gordon et al. (1993), one approach adds artificial evolution noise to the parameters and treats them as part of the state variables. Hence the original particle filters can be literally applied to models with unknown parameters. The main drawback of this approach is that the artificial noise would reduce the precision of parameter estimates and result in far too diffusive posteriors relative to the theoretical ones for the parameters.

In the context of structural break models, a similar idea of treating possibly time-varying parameters as state variables has appeared in Chopin (2007). Since the parameters remain constant with positive probability, the usual particle filters are highly inefficient and remedies such as local MCMC moves Gilks and Berzuini (2001) are introduced in order to artificially rejuvenate the parameter particles. This approach is generally not available when structural breaks affect only a subset of parameters as the conditional posterior for the constant parameters depends on the whole history of latent break points and the full sample of data.[8]

An alternative approach is proposed in Storvik (2002) on the basis that the posterior distribution of parameters $p(\theta|s_t, ..., s_1, D_t)$ is analytically tractable and depends on the observed data and latent state variables only through a vector of sufficient statistics. This requirement of sufficient statistics limits its applicability in certain models. A related approach is discussed in Johannes and Polson (2006).

In this chapter we follow Liu and West (2001) to incorporate parameter learning. It is based on the Bayes decomposition of the posterior density

$$p(s_{t+1}, \theta|D_{t+1}) \propto p(y_{t+1}|s_{t+1}, \theta, D_t)p(s_{t+1}|\theta, D_t)p(\theta|D_t)$$

and uses a kernel smoothing method with shrinkage to approximate $p(\theta|D_t)$. Specifically, the posterior density $p(\theta|D_t)$ is approximated by a mixture of multivariate normals

$$p(\theta|D_t) \approx \sum_{i=1}^{N} w_t^{(i)} N(\theta|a\theta_t^{(i)} + (1-a)\overline{\theta}_t, b^2 V_t)$$

where $\overline{\theta}_t = \sum_{i=1}^{N} w_t^{(i)}\theta_t^{(i)}$ and $V_t = \sum_{i=1}^{N} w_t^{(i)}(\theta_t^{(i)} - \overline{\theta}_t)(\theta_t^{(i)} - \overline{\theta}_t)'$, and $N(\theta|.,.)$ is the multivariate normal pdf. The constants a and b measure the extent of the shrinkage and are determined via a discount factor $\delta \in (0, 1)$ as $a = \sqrt{1-b^2}$ and $b^2 = 1 - [(3\delta - 1)/2\delta]^2$.

[8]The path dependence will also affect a GARCH model in which all parameters change from a break. Then local MCMC moves would require the history of states to perform sampling. One possible solution could be to reset the volatility process after a break so that the conditional variance starts up from a preset value or the implied unconditional value and is not a function of past data. A drawback is that there may be a drop in the precision of the conditional variance.

Liu and West (2001) shows that this kernel smoothing approach connects to the artificial evolution approach but removes the problem of information loss over time via shrinkage.

Conditional on samples of θ drawn from the mixture, the usual particle filters can be applied to estimate the state variables. As shown in Liu and West (2001), the kernel smoothing approach combined with an efficient particle filter such as the APF produces rather efficient estimates.[9] It also has the additional benefit of small extra computation cost.

2.3.3 A Robust Algorithm for Structural Break Models

We combine the APF with the kernel smoothing approach to design a sequential Monte Carlo algorithm for the general structural break model. The state at time t, s_t, will equal the number of states that have actually appeared in the studied time series up to time t. In contrast to the MCMC method of Chib (1998), this approach does not enforce that all states be visited.[10] We follow Chopin (2007), and Chopin and Pelgrin (2004), and specify an upper bound $\overline{K}$ on the number of states to facilitate computation. As long as the upper bound on the number of states is large enough we can jointly estimate the model parameters and the number of structural breaks at each point in time. Therefore, the problem of determining the number of states is automatically solved by sequentially learning about the state variables over time. We consider this a major benefit of sequential Monte Carlo methods for structural break models.

The parameters θ should be reparameterized to take values on the real line since they will be sampled through a mixture of normal kernels. For example, the probability p_{ij} can be reparameterized as $\log(\frac{p_{ij}}{1-p_{ij}})$. To simplify notation, we will use θ to denote the transformed parameters as well in the following discussion.

[9] Casarin and Marin (2007) find this method to be the best among several other alternatives for estimation of fixed parameters and states for a stochastic volatility model.

[10] The MCMC method in Chib (1998) samples the state variables backward from the end of sample and assumes that the final state is $\overline{K}$ and the first state in the sample is 1 so that all states are visited.

To initialize the algorithm, we need to specify a upper bound $\overline{K}$ for the possible number of states in the studied dataset. The only requirement for $\overline{K}$ is that it should be larger than the estimated number of states so that the number of states is freely determined. Let $\{\theta_{k,t}^{(i)}\}_{i=1}^{N}$, $k = 1, 2, ..., \overline{K}$, denote the particles of the parameters in state k and let $\theta_t^{(i)} \equiv \{\theta_{1,t}^{(i)}, ..., \theta_{\overline{K},t}^{(i)}\}$. Let $p_t^{(i)}$ denote the transition probability components of $\theta_t^{(i)}$. Note the subscript t indicates that the parameter values are learned at time t, and not necessarily time-varying. Initial particles and weights are set as $s_1^{(i)} = 1, w_1^{(i)} = \frac{1}{N}$, for $i = 1, 2, ..., N$ and parameters are drawn from the prior. The prior $\{\theta_1^{(i)}\}_{i=1}^{N}$ will depend on the specific model.

Our initial algorithm which employed the standard multinomial resampling performed poorly in practice, delivering rather unstable estimates of the parameters and states over repeated runs. The discrepancy of estimates between multiple runs of this algorithm are usually unacceptably large. An efficient sampling scheme for drawing the auxiliary indices is necessary to reduce the Monte Carlo variation and stabilize the estimates.

To stabilize estimates over multiple runs we use stratified sampling of Carpenter, Clifford, and Fearnhead (1999).[11] To produce a new sample of size m from a population $\{x_t\}_{t=1}^{N}$ with weights $\{w_t\}_{t=1}^{N}$, stratified sampling first produces stratified uniform random variables $\{u_t\}_{t=1}^{m}$ by drawing $u_t \sim U(\frac{t-1}{m}, \frac{t}{m})$ independently. From each of these draws x_t is selected based on multinomial sampling. We find this method fast and effective in stabilizing estimates across different runs.

Given a set of particles and weights $\{\theta_t^{(i)}, s_t^{(i)}, w_t^{(i)}\}_{i=1}^{N} \sim p(s_t, \theta|D_t)$, the following is a general algorithm for structural break models following Liu and West (2001) with the addition of stratified sampling.

1. For $i = 1, 2, ..., N$, let r be the mode of $p(s_{t+1}|s_t^{(i)}, \theta_t^{(i)})$. Compute the weights for the auxiliary index $\widetilde{w}_{t+1}^{(i)} \propto p(y_{t+1}|D_t, a\theta_{r,t}^{(i)} + (1-a)\overline{\theta}_{r,t})w_t^{(i)}$, where $\overline{\theta}_{r,t} = \sum_{i=1}^{N} w_t^{(i)}\theta_{r,t}^{(i)}$.

[11] We also investigated residual sampling (Liu and Chen (1998)) and found this was effective in stabilizing the estimates. However, stratified sampling is both easy to program and can directly exploit the $O(N)$ computational speed of multinomial sampling.

2. Draw stratified uniform random variables $\{u_j\}_{j=1}^N$ with $u_j \sim U(\frac{j-1}{N}, \frac{j}{N})$.

3. Produce the auxiliary indices $\{r^i\}_{i=1}^N$, $r^i \in \{1, 2, ..., N\}$, by retaining N^i copies of i, $i = 1, 2, ..., N$, where N^i is the number of $\{u_j\}_{j=1}^N$ falling in the interval $(\sum_{j=0}^{i-1} \widetilde{w}_j, \sum_{j=0}^{i} \widetilde{w}_j]$ with $\widetilde{w}_0 = 0$.

4. For $i = 1, 2, ..., N$, sample $\theta_{t+1}^{(i)}$ from $N(m_t^{(r^i)}, b^2 V_t)$, where $m_t^{(r^i)} = a\theta_t^{(r^i)} + (1 - a)\overline{\theta}_t, V_t = \sum_{i=1}^N w_t^{(i)}(\theta_t^{(i)} - \overline{\theta}_t)(\theta_t^{(i)} - \overline{\theta}_t)'$.

5. For $i = 1, 2, ..., N$, sample $s_{t+1}^{(i)} \in \{1, 2, ..., \overline{K}\}$ from $p_{t+1}^{(i)}(s_{t+1}|s_t^{(r^i)})$.

6. Let $s = s_{t+1}^{(i)}$. Compute $w_{t+1}^{(i)} \propto p(y_{t+1}|D_t, \theta_{s,t+1}^{(i)})/p(y_{t+1}|D_t, a\theta_{r,t}^{(r^i)} + (1 - a)\overline{\theta}_{r,t})$.

This gives the sample $\{s_{t+1}^{(i)}, \theta_{t+1}^{(i)}, w_{t+1}^{(i)}\}_{i=1}^N \sim p(s_{t+1}, \theta|D_{t+1})$.

2.3.4 Predictive Likelihoods and Model Selection

The predictive likelihood $p(y_{t+1}|D_t)$ can be computed from

$$p(y_{t+1}|D_t) = \int\int\int p(y_{t+1}|s_{t+1}, \theta)p(s_{t+1}|s_t, \theta)p(s_t, \theta|D_t)ds_{t+1}ds_t d\theta. \tag{2.8}$$

Once a set of particles and weights $\{s_t^{(i)}, \theta_t^{(i)}, w_t^{(i)}\}_{i=1}^N \sim p(s_t, \theta|D_t)$ is available, one can compute the approximation

$$p(y_{t+1}|D_t) \approx \sum_{i=1}^N w_t^{(i)} \int p(y_{t+1}|s_{t+1}, \theta_t^{(i)})p(s_{t+1}|s_t^{(i)}, \theta_t^{(i)})ds_{t+1}$$

where

$$\begin{aligned}\int p(y_{t+1}|s_{t+1}, \theta_t^{(i)})p(s_{t+1}|s_t^{(i)}, \theta_t^{(i)})ds_{t+1} &= p(y_{t+1}|s_{t+1} = s_t^{(i)}, \theta_t^{(i)})p(s_{t+1} = s_t^{(i)}|s_t^{(i)}, \theta_t^{(i)}) \\ &+ p(y_{t+1}|s_{t+1} = s_t^{(i)} + 1, \theta_t^{(i)})p(s_{t+1} = s_t^{(i)} + 1|s_t^{(i)}, \theta_t^{(i)}).\end{aligned} \tag{2.9}$$

If the information set is $D_t = \{y_1, y_2, ..., y_t\}$, the marginal likelihoods can be computed sequentially via predictive likelihoods

$$p(y_1, ..., y_t) = p(y_1) \prod_{\tau=2}^t p(y_\tau|y_1, ..., y_{\tau-1}), \quad t = 2, ..., T$$

By construction, the marginal likelihood can be interpreted as a measure of the cumulative out-of-sample predictive power of the model under investigation. Sequential Bayes factors (the ratio of marginal likelihoods of two specifications) can be then used to conduct model selection. Let $BF_{AB} = \frac{p(y_1,...,y_T|Model\ A)}{p(y_1,...,y_T|Model\ B)}$. Kass and Raftery (1995) suggest interpreting the evidence for A as: not worth more than a bare mention if $0 \leq BF_{AB} < 3$; positive if $3 \leq BF_{AB} < 20$; strong if $20 \leq BF_{AB} < 150$; and very strong if $BF_{AB} \geq 150$.

2.4 Structural Break GARCH Models

As discussed above, the final state $s_T = k$, where $k \leq \overline{K}$, is determined by the data rather than being specified *a priori*. The state variable s_t evolves according to the transition probability matrix in Equation 2.1. The chain can either stay in the current state or go to the next one. The final state k is equal the number of in-sample states and the number of break points is $k - 1$.

Let $\theta_k = [c_k, \alpha, \beta, P]$ be the model parameters in state k and $\theta = \{\theta_1, ..., , \theta_{\overline{K}}\}$. The likelihood of y_t for the partial break model in Equation 2.2, when $\epsilon_t \sim NID(0, 1)$ is

$$\begin{aligned} p(y_t|D_{t-1}, s_t, \theta) &= p(y_t|D_{t-1}, \theta_{s_t}) && (2.10) \\ &= \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{y_t^2}{2\sigma_t^2}\right) && (2.11) \end{aligned}$$

where $\sigma_t^2 = c_{s_t} + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2$. We refer to this structural break specification as a partial SB-GARCH-N model since only the intercept of the conditional variance is subject to breaks.

Empirical studies often suggest fat tails in the distribution of asset returns, therefore, an alternative specification for the return innovation ϵ_t would be a student-t distribution with v degrees of freedom. Let SB-GARCH-t denote this model and if $\theta_k = [c_k, \alpha, \beta, v, P]$,

then the data density of y_t becomes

$$\begin{aligned} p(y_t|D_{t-1}, s_t, \theta) &= p(y_t|D_{t-1}, \theta_{s_t}) && (2.12) \\ &= \frac{\Gamma(\frac{v+1}{2})}{\sigma_t \Gamma(\frac{v}{2})\sqrt{\pi v}} \left(1 + \frac{y_t^2}{v\sigma_t^2}\right)^{-\frac{v+1}{2}} && (2.13) \end{aligned}$$

where $\sigma_t^2 = c_{s_t} + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$.

It is possible that all parameters of the volatility process, not just the unconditional variance, may be subject to structural breaks. So we also consider a full SB-GARCH model (SB-GARCH-N and SB-GARCH-t)

$$y_t = \sigma_t \epsilon_t \tag{2.14}$$

$$\sigma_t^2 = c_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t} \sigma_{t-1}^2 \tag{2.15}$$

where $\theta_k = [c_k, \alpha_k, \beta_k, P]$, or $\theta_k = [c_k, \alpha_k, \beta_k, v, P]$, depending on the specification of the innovation ϵ_t. The algorithm presented in Section 2.3.3 can be applied to estimate these models and compute their marginal likelihoods for model comparison purposes.

2.5 Simulation Evidence

To analyze the performance of the proposed algorithm, we simulated 3000 observations from the partial SB-GARCH-N model with 2 break points. The parameter values are as follows

$$c_1 = 0.2,\ c_2 = 0.6,\ c_3 = 0.1$$

$$\alpha = 0.1,\ \beta = 0.8$$

with the break points at 1000 and 2000. A plot of the simulated data is presented in Figure 2.1.

The size of particles is set to be 300,000 and the upper bound of possible states $\overline{K} = 5$. Following Liu and West (2001), the discount factor is set as $\delta = 0.99$. In the empirical

work we restrict $p_{ii} = p$ for parsimony. The priors are as follows

$$c_i \sim Gamma(1, 0.2), \quad i = 1, ..., \overline{K}$$

$$\alpha \sim Beta(1, 8), \ \beta \sim Beta(4, 1)$$

$$\log\left(\frac{p}{1-p}\right) \sim N(10, 1)$$

when using the student-t distribution for returns, the prior on the degrees of freedom is set as $v \sim \text{Gamma}(1.2, 30)$.

In our experiments, we find that a prior on the transition probability p close to one is critical for obtaining sensible results since estimates of states are sensitive to outliers.[12] A prior for $\log\left(\frac{p}{1-p}\right)$ centered on lower values (ie. 5) quickly results in all states being visited and being stuck in the final state $\overline{K}$ for the rest of the sample. Therefore, to avoid state saturation it is necessary to have an informative prior. As such our prior favors no breaks. On the other hand the priors for remaining parameters cover a wide range of values consistent with existing empirical studies on GARCH models.

All estimations in this chapter are performed using the GNU scientific library in C. Each filtering iteration takes about 2-3 seconds. The estimates of states are presented in the middle panel of Figure 2.1. The lower panel of Figure 2.1 provides the distribution of the state estimates (the filter $P(s_t|D_t)$) over time, which shows clearly the evolution of the state particles. It can be seen that the algorithm is able to successfully identify the number and locations of break points. There are a few spikes in the filter estimates but they are quickly corrected. Intuitively, these spikes are caused by outlier observations. The state particles in these "false" states are subsequently dominated by particles which stay in the "true" states as the algorithm updates information by using new observations and gives higher weights to these correct particles.

The full-sample parameter estimates are presented in Table 2.1 and their sequential posterior means and 95% credible intervals are plotted in Figure 2.2. For all parameters,

[12] A false estimate of states usually leads to poor estimates of parameters.

the posteriors collapse nicely to their true values. It is worthwhile to note that estimates of the parameters c_2 and c_3, which are specific to state 2 and 3 respectively, change sharply upon the break points with their credible intervals quickly shrinking afterwards. This pattern suggests that new information is correctly used and estimates are timely updated.

Figure 2.3 presents the filtered volatility estimates along with the volatilities from the true model. It can be seen that the estimates track the true volatilities closely. We also plot the sequence of the effective sample size of particles in Figure 2.4 to check for weight degeneracy. The effective sample size of particles with weights $\{w_t^{(i)}\}_{i=1}^N$ is computed as $1/\sum_{i=1}^N (w_t^{(i)})^2$, which equals N when all weights equal $1/N$ whereas more concentrated distributions of weights lead to smaller values of effective sample size. In the plot, there are sporadic drops in the effective sample size, usually around large outliers and break points, but they return to normal quickly afterwards, suggesting that weight degeneracy is minor in this study.

To see if the algorithm has the power to detect the true model when there are competing ones, we also estimate a partial SB-GARCH-t model, a full SB-GARCH-N model and a standard GARCH model with normal innovations for the simulated data. The results are presented in Table 2.1. The partial SB-GARCH-N model has the largest marginal likelihood, with the partial SB-GARCH-t model closely behind. The no-break GARCH model has a much smaller marginal likelihood than the structural break alternatives. We also plot the sequences of marginal likelihoods differences through time between the partial SB-GARCH-N model and its competitors in Figure 2.5. Based on these plots, the partial SB-GARCH-N model consistently outperforms the full SB-GARCH-N and no-break GARCH models whereas it does slightly better than the partial SB-GARCH-t model.[13]

[13]This is expected as the SB-GARCH-t model nests the true SB-GARCH-N model. The marginal likelihood imposes a penalty on this model from the additional parameter uncertainty of the degree of freedom, v.

2.6 Empirical Application

In this section, we apply the proposed algorithm to the daily NASDAQ composite returns from January 3, 1995 to December 29, 2006 (3022 observations). The data source is the Center for Research in Security Prices (CRSP). Table 2.2 provides the summary statistics of the data series. In estimation, returns are demeaned and scaled up by 100. The priors, number of particles, $\overline{K}$, a, b, and δ are the same as in the simulation experiments.

2.6.1 Estimation Results

To compare the performance of competing SB-GARCH models for the NASDAQ returns, we consider estimating 4 models: the partial SB-GARCH models with normal and t innovations and the full SB-GARCH models with normal and t innovations. To explore the importance of modeling structural breaks, we also fix $\overline{K} = 1$ and estimate two no-break GARCH models with normal and t innovations respectively. The estimated full-sample marginal likelihoods of the six models are presented in Table 2.3. Based on these estimates, the partial SB-GARCH-t model is the most favored while the versions in which all parameters change after a break have lower marginal likelihoods.

We first provide a comparison of the SB-GARCH-t models with the SB-GARCH-N models, which are the ones most commonly studied in the existing literature. The full-sample marginal likelihoods of Table 2.3 provide overwhelming evidence in favor of the specification with t-innovations: the t models have larger marginal likelihoods than the normal models in both the partial SB-GARCH and the no-break categories. Second each of the break models provides a large improvement over the no-break alternative. For instance, the log-Bayes factors are 22.99 (normal innovations) and 9.46 (t-innovations) in favor of the partial break model.

The sequentially filtered estimates of states and parameters of the partial SB-GARCH-N model are presented in Figures 2.6 and 2.7. This model identifies two break points

associated with a sustained increase in NASDAQ volatility lasting from late 1997 to early 2004.[14] The full-sample posterior means and 95% credible sets of parameter estimates are reported in Table 2.4. The resulting persistence estimate $\alpha + \beta = 0.966$ is lower than the estimate of 0.983 by the no-break GARCH model with normal innovations. The full SB-GARCH-N model has much smaller values of $\alpha + \beta$ in each regime. This finding is consistent with the existing studies of GARCH models that find lower persistence once structural breaks in the volatility process are taken into account.

The results are different when we fit the more flexible partial SB-GARCH-t model to the data. Figure 2.8 and 2.9 present the sequentially filtered estimates of states and parameters of this model. It identifies a similar pattern of breaks, but there is more uncertainty. Recall that our analysis addresses the real time filtering problem, and as such smoothed historical estimates of breaks may be more precise. By the end of the sample there is substantial probability that no break has occurred. For example, at the end of the sample $p(S_T = 1|D_T) = 0.671$, $p(S_T = 2|D_T) = 0.0002$, and $p(S_T = 3|D_T) = 0.321$. Compared to the specification with normal innovations this model has a built in robustness from the fat tailed t-innovations. A structural break can temporarily appear as tail observations and it may take many observations to disentangle these effects. This is consistent with Maheu and McCurdy (2009) who model the unconditional distribution of returns and find less breaks when fat-tailed innovations are used instead of normal innovations. Nevertheless, Bayes factors favor the structural break model even with the large amount of uncertainty about breaks.

Sequential parameter estimates appear in Figure 2.9. Early in the sample the degree of freedom parameter takes a large drop. The first break found in the SB-GARCH-N model (Oct. 27,1997) is classified as an increase in tail thickness by this model (v drops in Figure 2.9). It is not till closer to mid-sample that the evidence for a structural break

[14]The dates of the break points are October 27, 1997 and January 23, 2004 based on the mode of the filtering density $p(s_t|D_t)$.

increases.

In common with the other break models Table 2.4 shows that $\alpha + \beta$ is lower for the t-innovation specifications when break are allowed. We also find that by the end of the sample the degree of freedom parameter is larger, (27 versus 19 for the no break model) for the SB-GARCH-t model.

The filtered conditional standard deviations through time are displayed in Figure 2.10 for 2 alternative models. Although the volatility estimates are broadly similar, panel B which displays their difference, show that the no break GARCH-N tend to produce larger estimates and is persistently larger near the end of the sample. As was mentioned above, model estimates may attribute higher persistence to volatility when breaks are not modeled.

Figures 2.11 and 2.12 provide information on the predictive value of modeling breaks. The first figure displays the cumulative log-Bayes factor in favor of the partial SB-GARCH-N against the no break alternative. There are some minor gains mid-sample but the real improvements occur at the latter part of the sample. Here the improvements in the predictive densities are consistent and ongoing. Similarly for the partial SB-GARCH-t model, the gains come at the end of the sample. In the middle of the sample there are some penalties incurred from the increased model complexity of the break specification. There is always a trade-off between modeling the structural change and the resulting increase in parameter uncertainty around break points.

2.6.2 Robustness

To investigate weight degeneracy, we plot the sequence of survival rates for the particles in Figure 2.13. This estimate is the number of particles that survive in Step 3 of the stratified sampling version of the particle filter in Section 2.3.3 to the next iteration and is computed as (number of unique indices in $\{r^i\}_{i=1}^N)/N$. We can conclude from this plot that there is no evidence of persistent degeneracy. In addition, we found the effective

sample size (not shown) supported this conclusion and showed minor transient drops that are close to the full sample size of particles.

We experimented running the programs under different seeds of the random number generator to check the stability of estimates. The difference between parameter estimates across runs is generally less than 0.01 while the difference between estimated marginal likelihoods is less than 1. We found that 100,000 or more particles produced reliable results while less than this could be unstable across repeated runs.

Table 2.5 reports the sensitivity of the marginal likelihood for the partial SB-GARCH-t model for different priors. In each case, a change is made to the benchmark prior and this is listed in the first column of the table. Although there are some changes in parameter estimates and state inference, the improved predictions this model provides are robust to different priors.

A possibility is that outliers in returns may be identified as permanent structural breaks. To investigate the effect of including jumps we consider the following partial break GARCH-N model with jumps

$$\begin{aligned} y_t &= \sigma_t \epsilon_t + J_t \eta_t, \;\; \epsilon_t \sim N(0,1) && (2.16) \\ \sigma_t^2 &= c_{s_t} + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 && (2.17) \\ p(J_t = 0) &= q, \; p(J_t = 1) = 1 - q, \;\; \eta_t \sim N(0, h). && (2.18) \end{aligned}$$

Estimating the SB-GARCH models with jumps can be done by augmenting the latent states with the jump indicator J_t and applying the algorithm in Section 3 with straightforward extensions. Using a prior probability of no jump $q \sim Beta(40, 2)$, and jump size variance $h \sim Gamma(2, 1)$, while keeping priors of other parameters unchanged, we obtain similar estimates as in the SB-GARCH-t model. Model estimates appear in Table 2.4 while state and jump estimates are plotted in Figure 2.14. In particular, the break points identified by the SB-GARCH-N model with jumps are close to those by the SB-GARCH-t model. The marginal likelihood of the SB-GARCH-N model with jumps

is larger than that of the SB-GARCH-N model but slightly lower than the SB-GARCH-t model. We also notice that the parameter estimates of the SB-GARCH-N model with jumps are similar to those of the SB-GARCH-N model. We conclude that it is important to use a flexible model for daily returns, failure to do so may results in false identification of structural change.

Finally, we note that modeling jumps generally tends to reduce the effect of temporary outliers and hence enables lowering the strong prior on the transition probabilities of states. For example, using a prior of the transition probability $\log\left(\frac{p}{1-p}\right) \sim N(8.5, 1)$ produces similar results.

2.7 Conclusion

This chapter proposes a sequential Monte Carlo filtering algorithm to estimate GARCH models subject to structural breaks. There are several notable features of the proposed algorithm: the number of breaks is estimated simultaneously with other model parameters and states in a single run; the estimates of parameters and states are fast and efficiently updated once new observations become available; and by focusing on the sequential filtering problem the path dependence that structural breaks induce in GARCH models does not cause any problems for estimation.

Simulation examples show that the algorithm is able to perform accurate sequential inference. Our empirical application underscores the importance of model assumptions when investigating breaks. A model with normal return innovations result in strong evidence of breaks; while more flexible return distributions such as t-innovations or adding jumps to the model still favor breaks but indicate much more uncertainty regarding the time and impact of them. We also find that the partial structural break specification delivers better performance than the full structural break specification in which all parameters change from a break.

Table 2.1: Parameter Estimates for Simulated Data

	True Value	Partial SB-GARCH-N	Partial SB-GARCH-t	Full SB-GARCH-N	GARCH-N
c_1	0.2	0.167	0.169	0.229	0.088
		(0.102,0.259)	(0.103,0.262)	(0.076,0.541)	(0.064,0.117)
c_2	0.6	0.580	0.573	0.480	
		(0.333,0.946)	(0.331,0.922)	(0.242,0.852)	
c_3	0.1	0.098	0.098	0.160	
		(0.055,0.162)	(0.055,0.159)	(0.071,0.306)	
α_1	0.1	0.069	0.066	0.034	0.098
		(0.039,0.112)	(0.039,0.103)	(0.006,0.108)	(0.069,0.135)
α_2				0.071	
				(0.036,0.124)	
α_3				0.072	
				(0.029,0.145)	
β_1	0.8	0.834	0.829	0.821	0.875
		(0.751,0.898)	(0.748,0.893)	(0.617,0.941)	(0.834,0.908)
β_2				0.846	
				(0.757,0.912)	
β_3				0.755	
				(0.582,0.884)	
v			42.734		
			(21.239,77.134)		
LML		-5463.023	-5464.364	-5466.228	-5507.791

This table reports the full-sample posterior means of parameters for simulated data. Numbers in parenthesis are 95% credible sets. LML stands for log marginal likelihood.

Table 2.2: Summary Statistics of Daily NASDAQ Returns

Mean	Std. Deviation	Skewness	Kurtosis	Min.	Max.
0.054	1.715	0.174	7.936	-9.81	14.27

This table reports the summary statistics of NASDAQ return series from Jan 03,1995 to Dec 29,2006.

Table 2.3: Marginal Likelihoods

	Partial SB-GARCH-N	Partial SB-GARCH-t	Full SB-GARCH-N	Full SB-GARCH-t
LML	-5231.25	-5225.46	-5228.40	-5228.41
	GARCH-N	GARCH-t	Partial SB-GARCH-N-Jump	
LML	-5254.24	-5234.92	-5227.26	

This table reports the full-sample log marginal likelihoods for NASDAQ return series from Jan 03,1995 to Dec 29,2006. LML stands for log marginal likelihood.

Table 2.4: Parameter Estimates for NASDAQ Returns

	Partial SB-GARCH-N	Partial SB-GARCH-t	GARCH-N	GARCH-t	Partial SB-GARCH-N-Jump
c_1	0.070	0.043	0.052	0.033	0.048
	(0.038,0.113)	(0.027,0.067)	(0.032,0.071)	(0.023,0.0448)	(0.022,0.082)
c_2	0.171	0.211			0.231
	(0.090,0.303)	(0.097,0.391)			(0.115,0.417)
c_3	0.055	0.171			
	(0.006,0.324)	(0.006,0.947)			
α	0.074	0.062	0.095	0.070	0.064
	(0.054,0.098)	(0.043,0.085)	(0.072,0.124)	(0.044,0.088)	(0.046,0.090)
β	0.892	0.890	0.888	0.909	0.875
	(0.851,0.924)	(0.848,0.924)	(0.859,0.912)	(0.835,0.908)	(0.846,0.919)
v		27.190		19.268	
		(16.021,43.337)		(12.797,27.874)	
q					0.959
					(0.899,0.988)
h					2.149
					(0.695,5.135)

This table reports the full-sample posterior means of parameters for NASDAQ return series from Jan 03,1995 to Dec 29,2006. Numbers in parenthesis are 95% credible sets.

Table 2.5: Marginal Likelihood Estimates for Different Priors: Partial SB-GARCH-t

Prior	LML
benchmark	-5225.46
$c_i \sim Gamma(1, 0.4)$	-5227.18
$\alpha \sim U(0,1),\ \beta \sim U(0,1)$	-5226.46
$\log(p_{ii}/(1-p_{ii})) \sim N(8,1)$	-5220.04
$\overline{K} = 7$	-5225.91
$\nu \sim Gamma(1.2, 10)$	-5224.41

This table reports the log-marginal likelihood (LML) for the partial SB-GARCH-t model based on the benchmark prior in Section 2.5 with any changes listed in the first column.

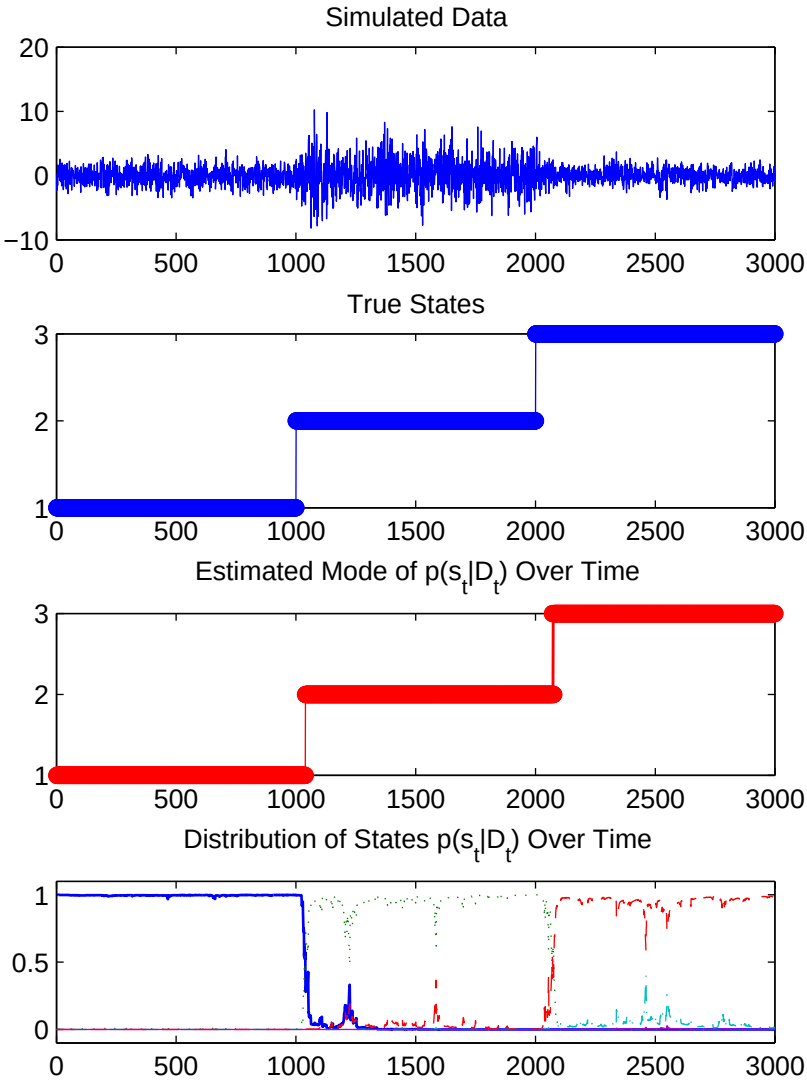

Figure 2.1: State Estimates of the Partial SB-GARCH-N Model, Simulated Data. For the distribution of states, bold solid: $p(s_t = 1|D_t)$, dot: $p(s_t = 2|D_t)$, dash: $p(s_t = 3|D_t)$, dash-dot: $p(s_t = 4|D_t)$, solid: $p(s_t = 5|D_t)$.

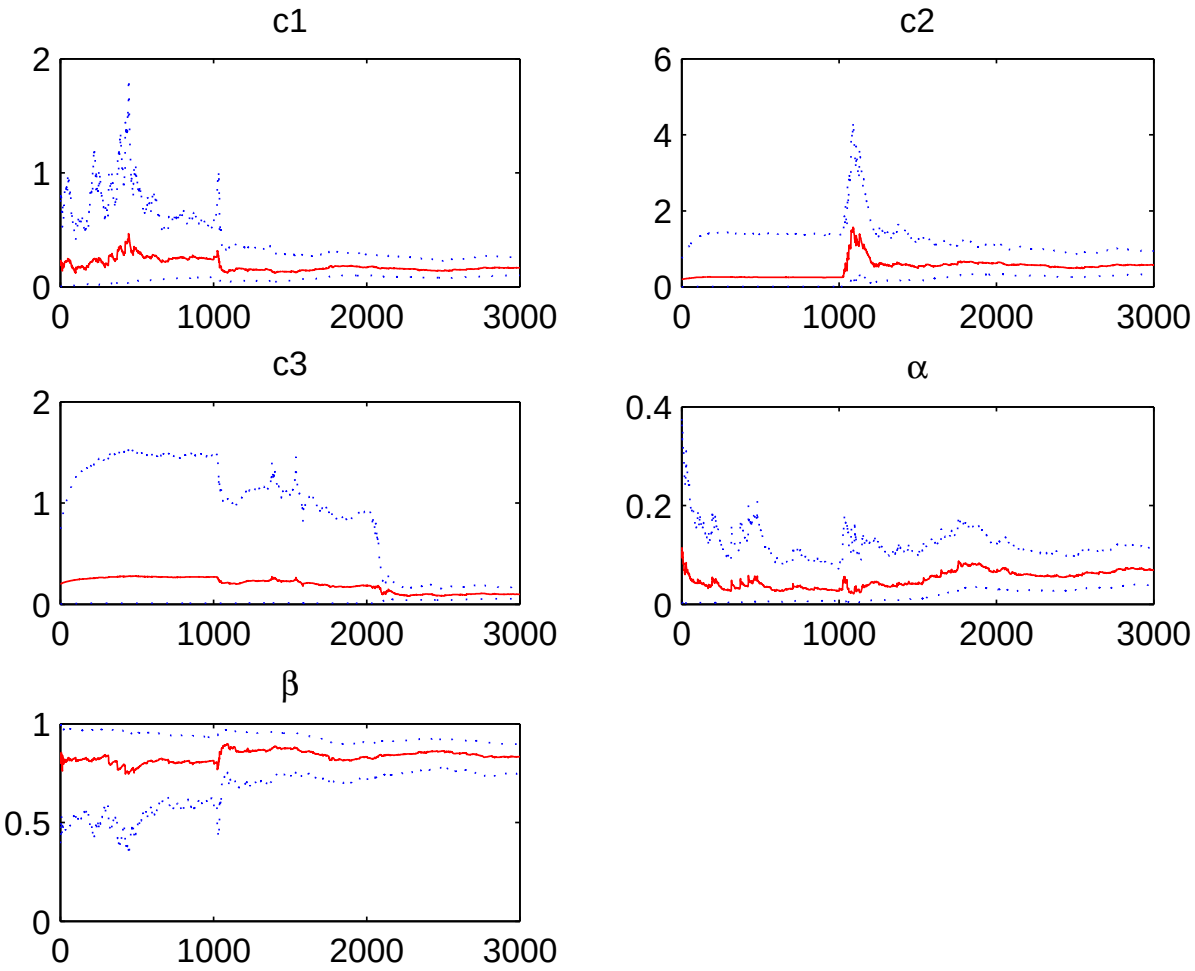

Figure 2.2: Sequential Parameters Estimates of the Partial SB-GARCH-N Model, Simulated Data.

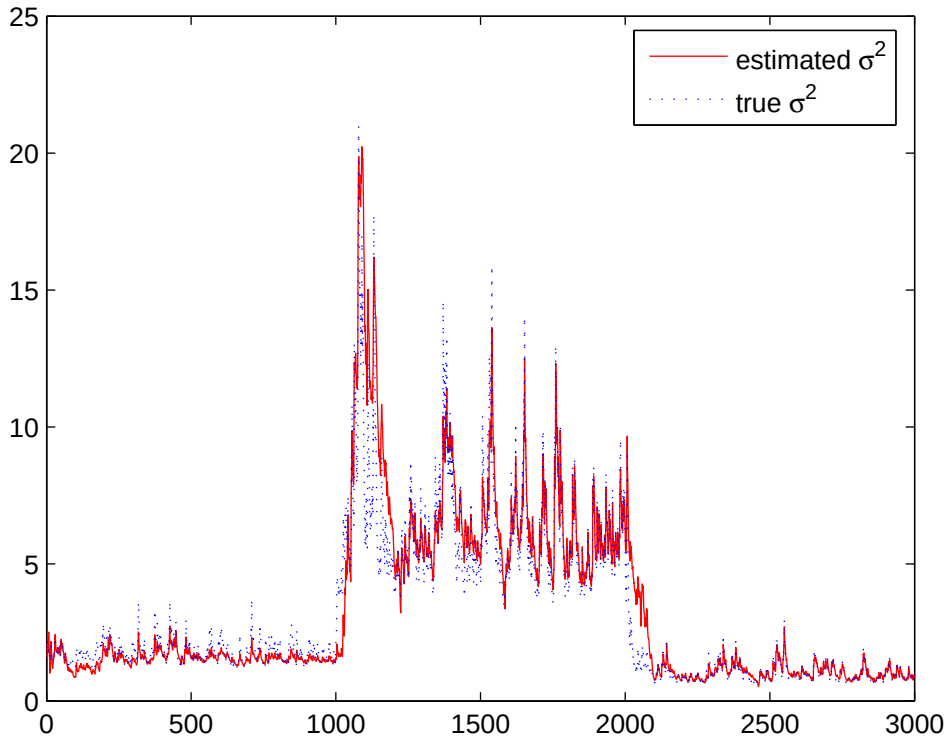

Figure 2.3: Volatility Estimates of the Partial SB-GARCH-N Model, Simulated Data.

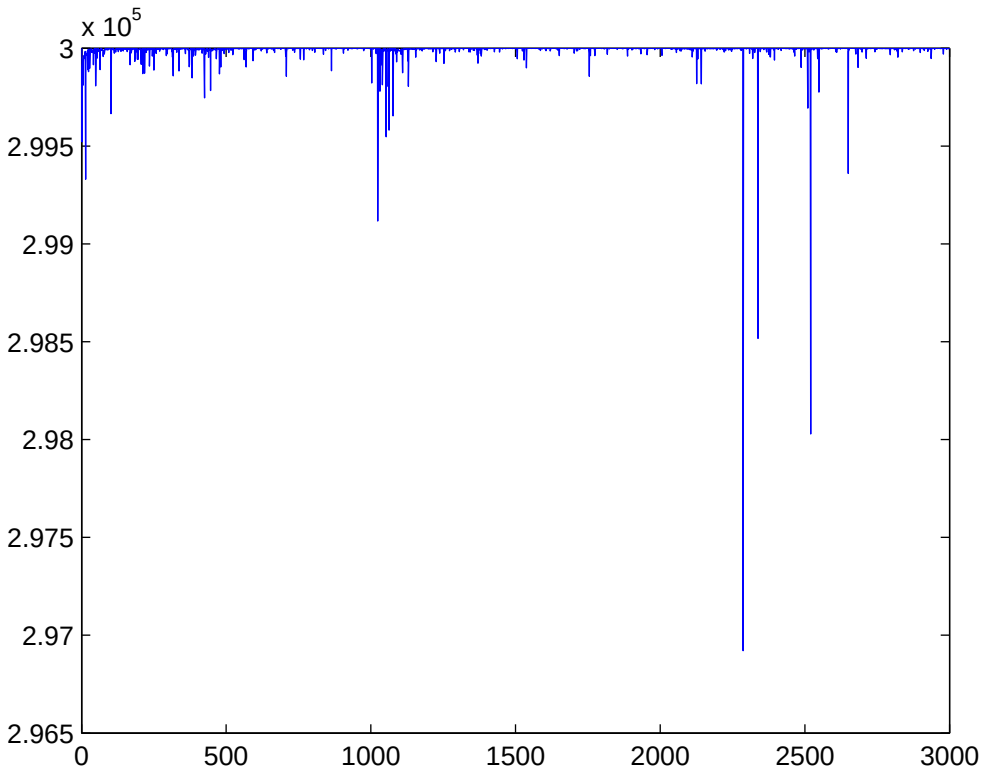

Figure 2.4: Effective Sample Size of Particles for the Partial SB-GARCH-N Model, Simulated Data

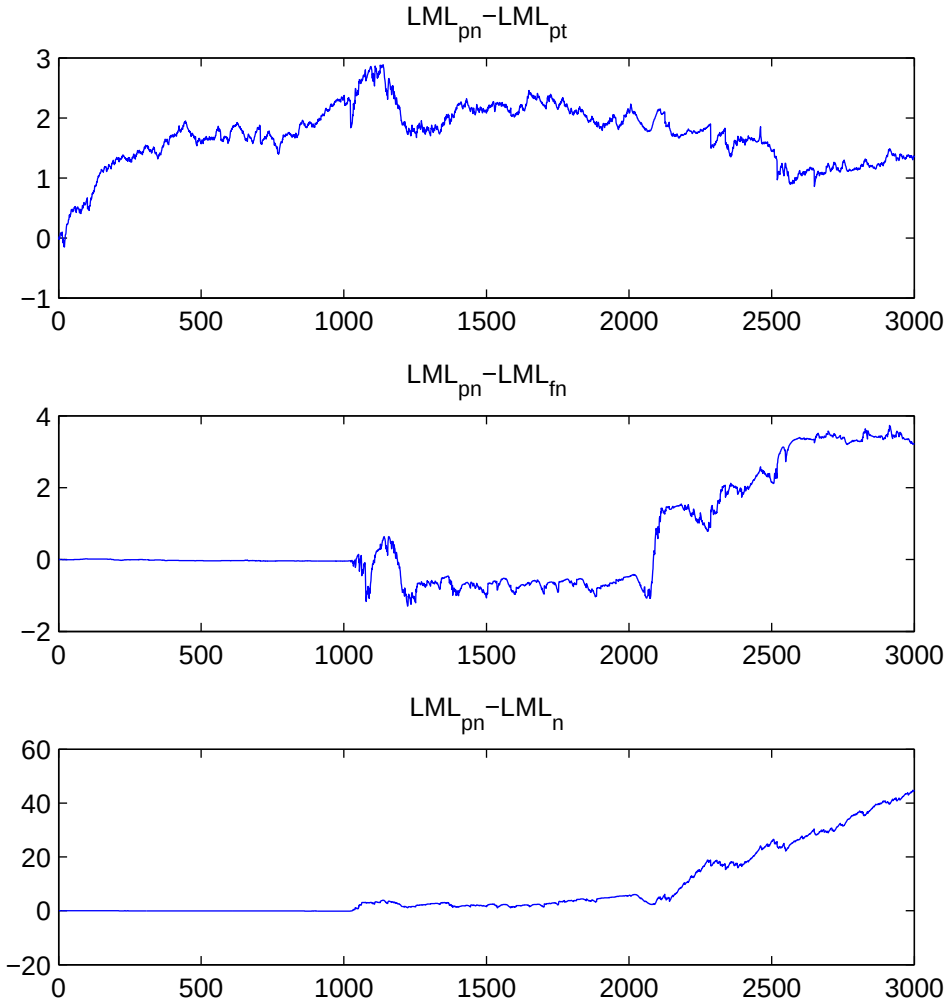

Figure 2.5: Cumulative Log-Bayes Factors for Partial SB-GARCH-N v.s. Misspecified Models, Simulated Data. LML_{pn} is the log marginal likelihood of the partial SB-GARCH-N model. LML_{pt} is the log marginal likelihood of the partial SB-GARCH-t model. LML_{fn} is the log marginal likelihood of the full SB-GARCH-N model. LML_n is the log marginal likelihood of the no-break GARCH-N model.

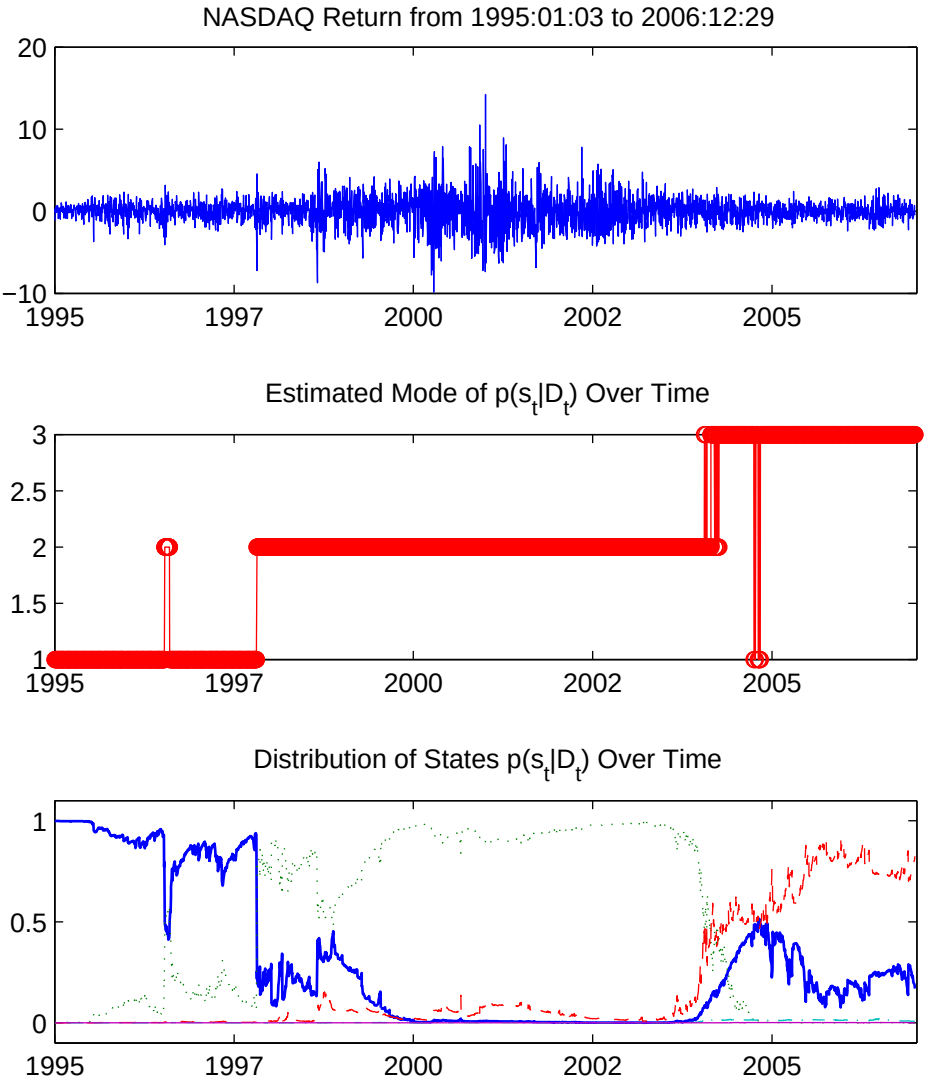

Figure 2.6: State Variable Estimates of the Partial SB-GARCH-N Model, NASDAQ Data. For the distribution of states, bold solid: $p(s_t = 1|D_t)$, dot: $p(s_t = 2|D_t)$, dash: $p(s_t = 3|D_t)$, dash-dot: $p(s_t = 4|D_t)$, solid: $p(s_t = 5|D_t)$.

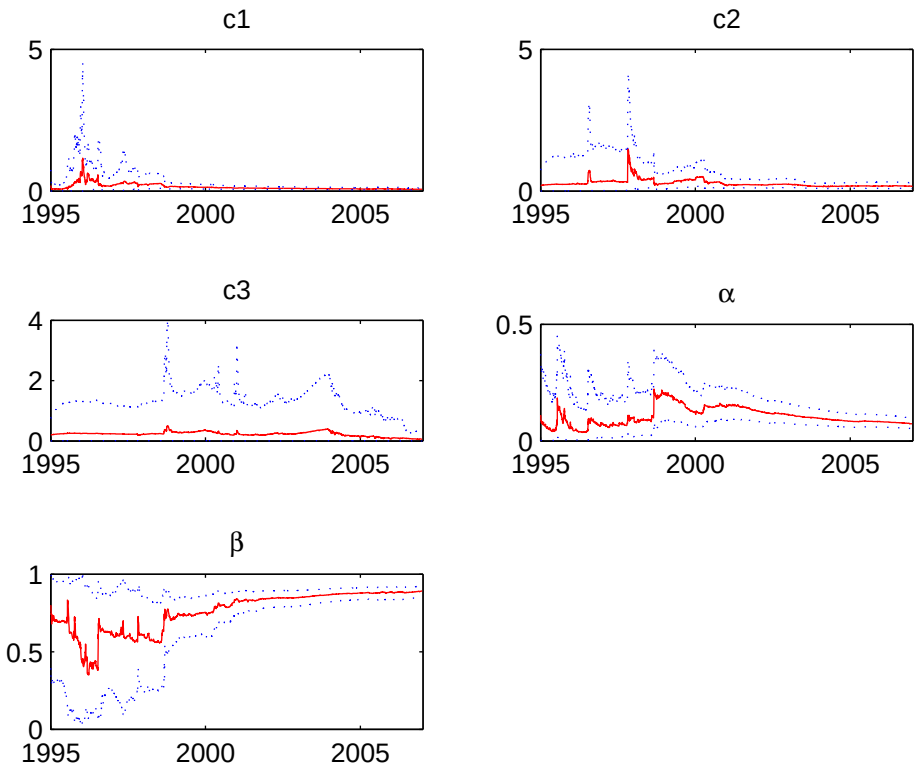

Figure 2.7: Parameters Estimates of the Partial SB-GARCH-N Model, NASDAQ Data

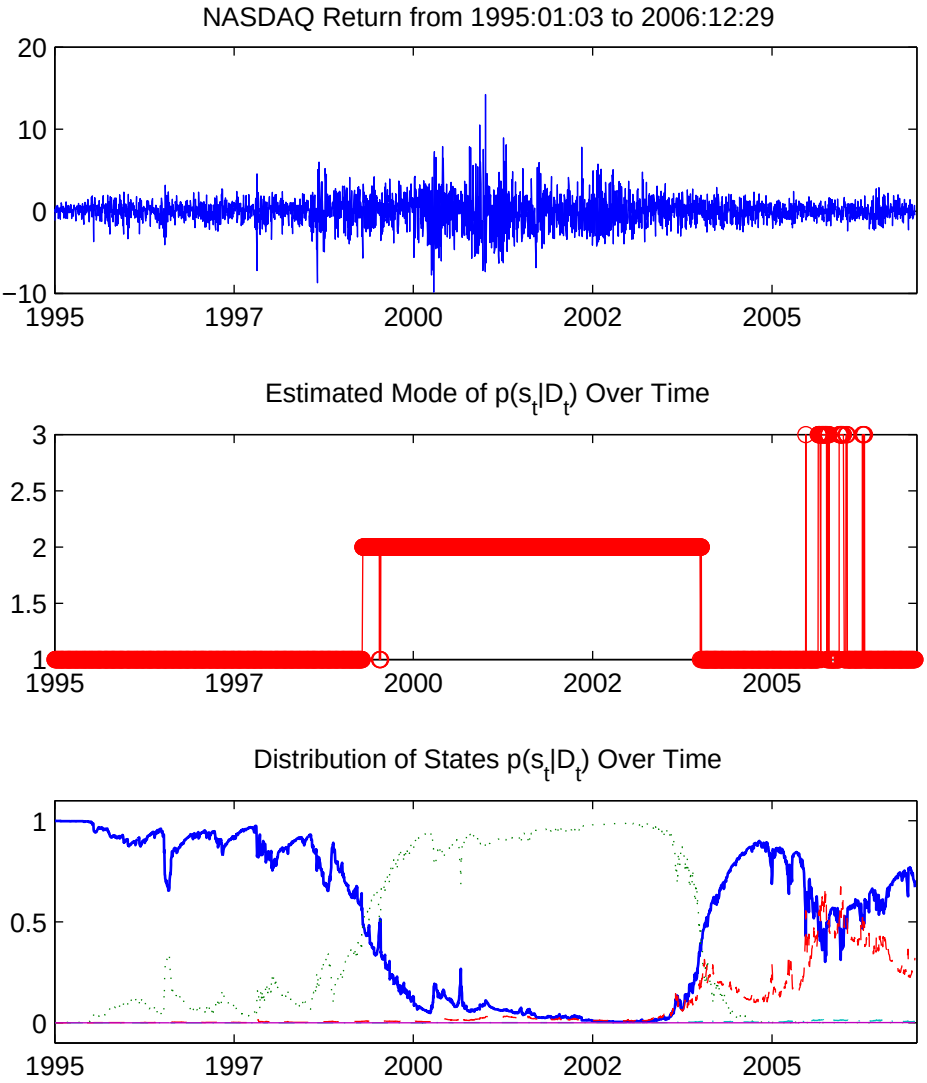

Figure 2.8: State Variable Estimates of the Partial SB-GARCH-t Model, NASDAQ Data. For the distribution of states, bold solid: $p(s_t = 1|D_t)$, dot: $p(s_t = 2|D_t)$, dash: $p(s_t = 3|D_t)$, dash-dot: $p(s_t = 4|D_t)$, solid: $p(s_t = 5|D_t)$.

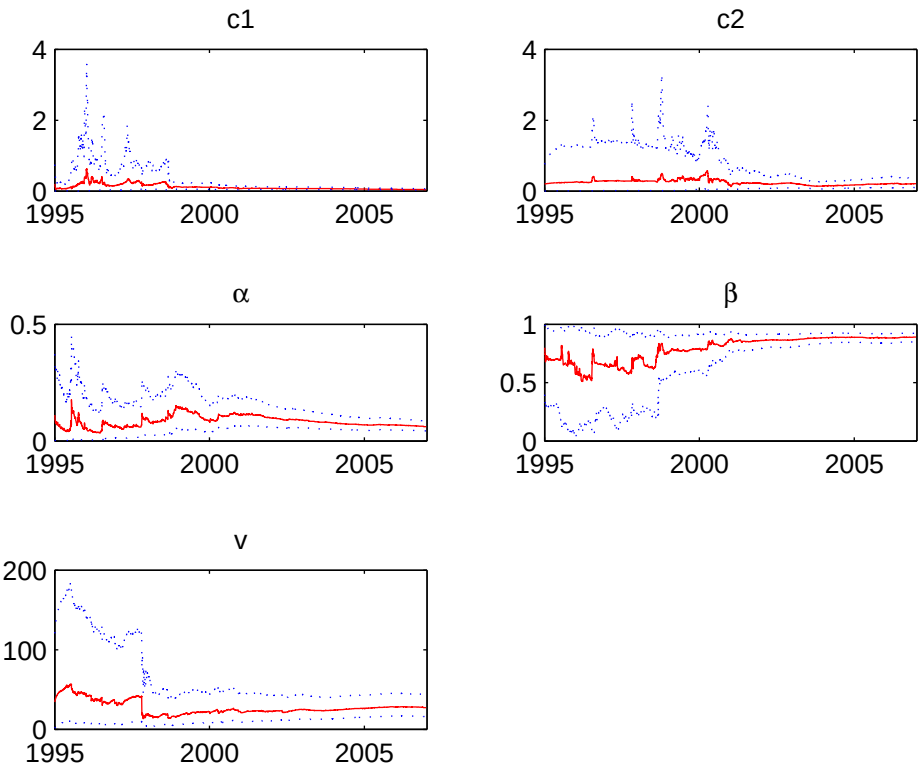

Figure 2.9: Parameters Estimates of the Partial SB-GARCH-t Model, NASDAQ Data

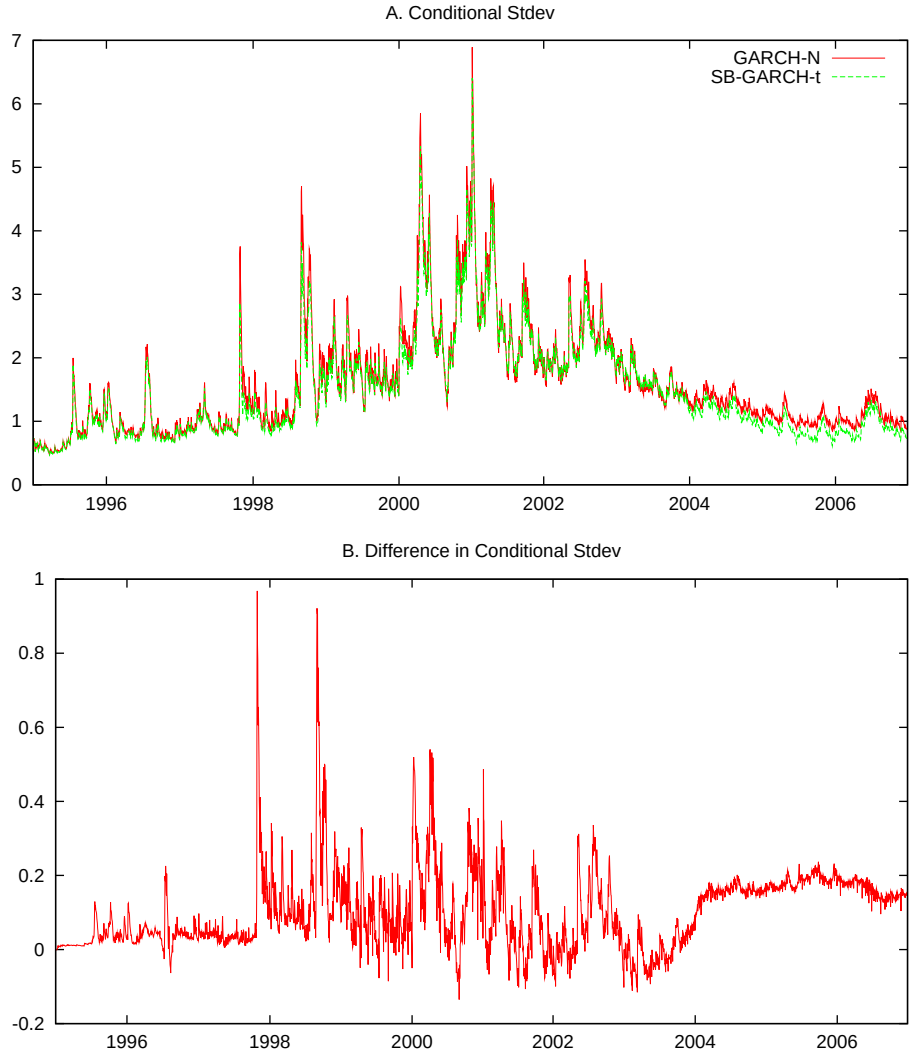

Figure 2.10: Conditional Standard Deviations: A. No-break GARCH-N v.s. partial SB-GARCH-t, and B. their difference

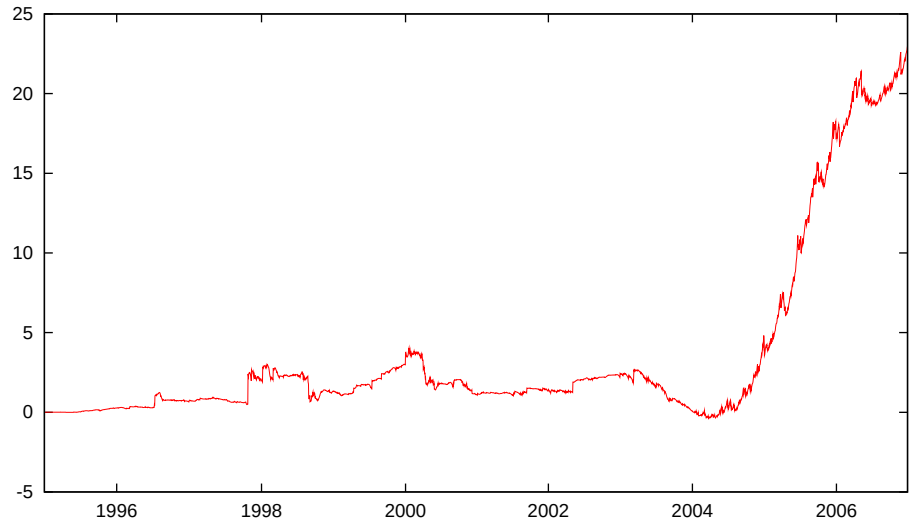

Figure 2.11: Cumulative Log-Bayes Factor: Partial SB-GARCH-N vs No Break GARCH-N

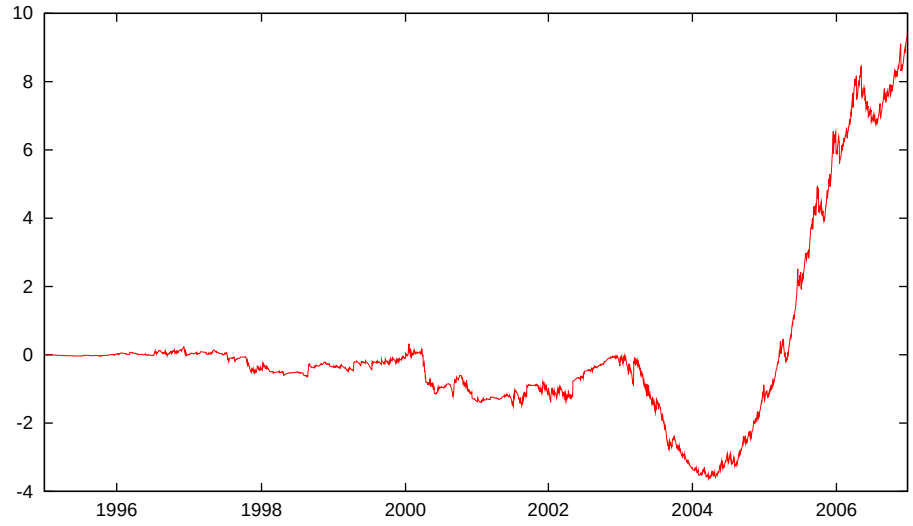

Figure 2.12: Cumulative Log-Bayes Factor: Partial SB-GARCH-t vs No Break GARCH-t

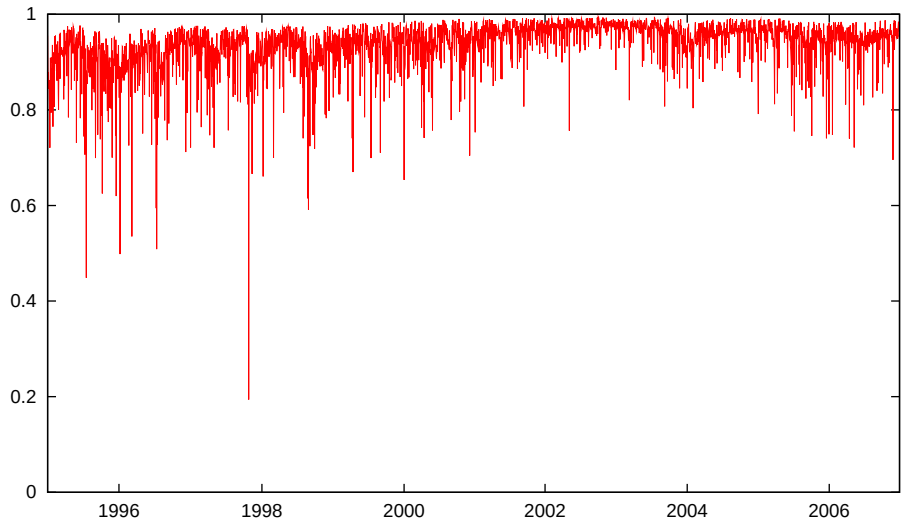

Figure 2.13: Survival Rate of Particles for the Partial SB-GARCH-t Model, NASDAQ Data

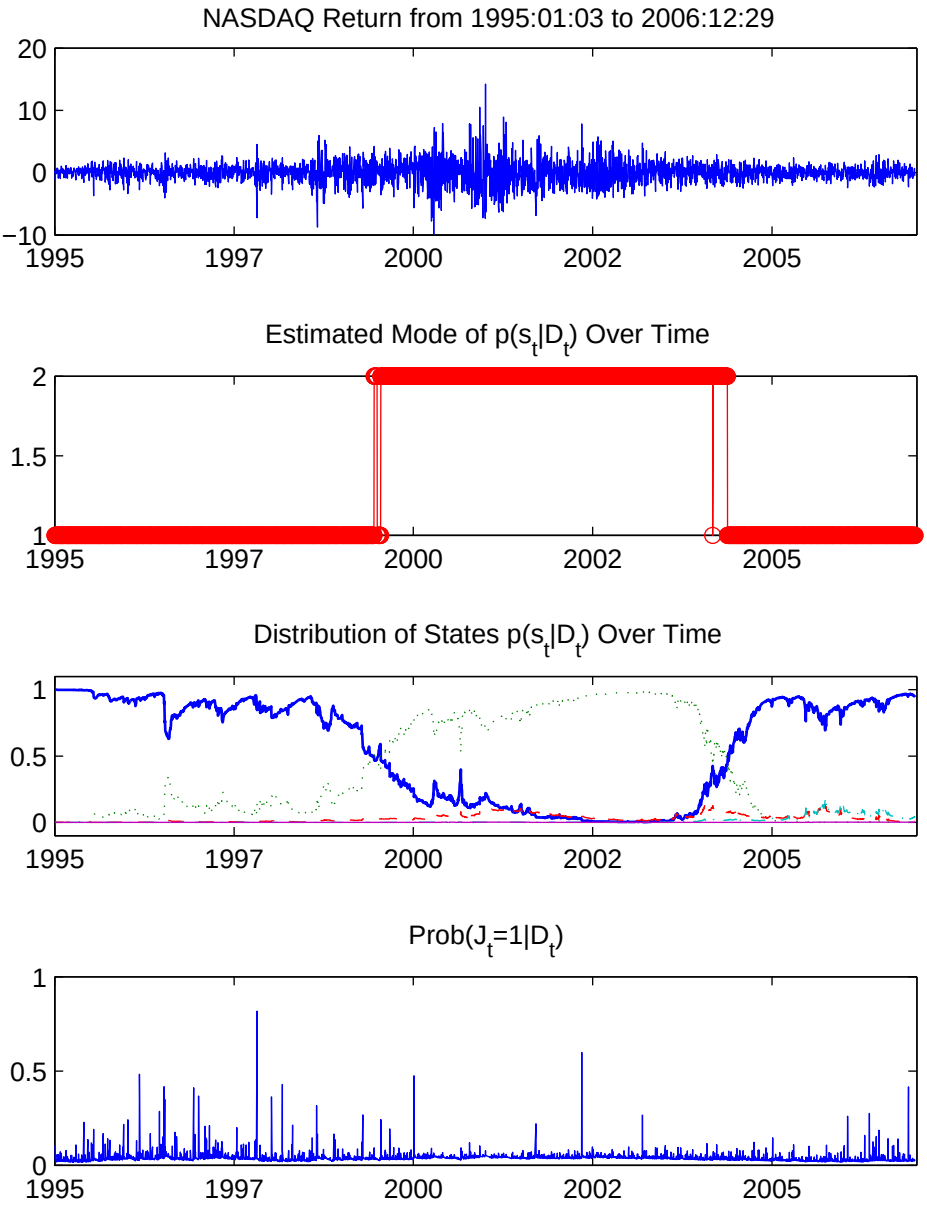

Figure 2.14: State Variable Estimates of the Partial SB-GARCH-N Model with Jumps, NASDAQ Data. For the distribution of states, bold solid: $p(s_t = 1|D_t)$, dot: $p(s_t = 2|D_t)$, dash: $p(s_t = 3|D_t)$, dash-dot: $p(s_t = 4|D_t)$, solid: $p(s_t = 5|D_t)$.

Chapter 3

A Tilt Stochastic Volatility Model with Leverage Effect

3.1 Introduction

The stochastic volatility (hereafter SV) models introduced by Taylor (1986) have attracted considerable attention in the recent econometric and finance literature because of their greater flexibility over the conventional GARCH family models (see, for example, Geweke (1994), Fridman and Harris (1998) and Kim, Shephard, and Chib (1998)) and their appeal as discrete-time approximations of the diffusion volatility processes used extensively in theoretical works, e.g. Hull and White (1987), Heston (1993) and Ball and Roma (1994). Ghysels, Harvey, and Renault (1996) and Shephard (2005) provide reviews of this class of models.

The basic SV (BSV) model assumes that the return and volatility innovations are uncorrelated, which is evidently inconsistent with the well documented leverage effect, that is, stock returns are negatively correlated with stock return volatility and negative returns are associated with larger impact on volatility than equally large positive returns (e.g. Black (1976), Christie (1982) and French, Schwert, and Stambaugh (1987)). Al-

though in the GARCH literature, there have been a variety of models which incorporate the leverage effect, introducing it into the SV model has proven to be more complicated, usually limited by the relative inflexibility of the estimation procedures.

Recent studies of the SV model with leverage effect (hereafter SVL) include Harvey and Shephard (1996) and Hansson and Hordahl (2005), which use the Quasi Maximum Likelihood approach for estimation, and Jacquier, Polson, and Rossi (2004) and Omori, Chib, Shephard, and Nakajima (2007), which develop Bayesian Markov chain Monte Carlo methods for estimation.[1] A common feature of these studies is that a single parameter, the correlation coefficient between the return and log volatility innovations, is used to capture the leverage effect and the asymmetric interaction between return and volatility is not explicitly modeled. The "leverage" effect arises in their models simply because of the device of modeling the natural logarithm of volatility: symmetric correlation between return and log volatility implies asymmetric correlation between return and volatility. But the degree of asymmetry is rigidly limited by the logarithmic functional form. As suggested by the numerous studies estimating the exponential GARCH (EGARCH) model of Nelson (1991), this natural device alone may be insufficient to capture the magnitude of the asymmetric correlation between return and volatility.

This chapter proposes a new SV model with leverage effect called the tilt stochastic volatility (TSV) model which explicitly models the asymmetric correlation between return and volatility innovations and nests the BSV model, the SVL model and the EGARCH model as special cases. The degree of asymmetry between the return and volatility innovations in the TSV model is determined by a free parameter and is estimated from data. Thus the leverage effect of stock returns is studied in a unified and flexible framework. The model is estimated by the Maximum Likelihood (ML) method based upon the Efficient Importance Sampling (EIS) procedure of Richard and Zhang (2007), which is a highly generic Monte Carlo technique for evaluating large dimensional

[1]See Yu (2005) for a discussion of leverage effect in stochastic volatility models.

integrals. The Maximum Likelihood (ML)-EIS estimator of the model is shown to work well in simulation studies. Applied to daily CRSP return data, I find that estimates of the extensions of the TSV model are significant. The estimated leverage effect from the TSV model is markedly different from the SVL and EGARCH models. In particular, the SVL model underestimates the degree of asymmetry of the correlation between return and volatility innovations while the EGARCH model overestimates it. Volatility estimates from the TSV model are more accurate than those from the BSV and SVL models. Compared with the EGARCH model, the TSV estimates of volatility track variations in volatility more closely but have larger finite sample bias.

The remainder of the chapter is organized as follows. Section 3.2 presents the TSV model and compares it with the existing volatility models. Section 3.3 describes in detail the estimation procedure. Section 3.4 provides a simulation study investigating the performance of the ML-EIS estimator for the TSV model. Section 3.5 applies the model to the US stock return data. Section 3.6 concludes.

3.2 The Model

Let r_t be the continuously compounded stock returns. The SV model takes the form

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ \log(\sigma_{t+1}^2) &= \gamma + \delta \log(\sigma_t^2) + \xi_t, \quad t = 1, 2, ..., n \end{aligned} \tag{3.1}$$

where $y_t \equiv r_t - E_{t-1}(r_t)$ is the demeaned return, σ_t^2 is the unobserved volatility of returns, ϵ_t is the return innovation with zero mean and unit variance, ξ_t is the innovation to log volatility with zero mean and the variance of σ_ξ^2.

For daily stock returns, a convention in the literature of SV models is to treat $E_{t-1}(r_t)$ as a constant or a simple linear function of past returns so to focus attention on the volatility process. To simplify notation, we will refer to y_t as return in the rest of the chapter.

To discuss the leverage effect in SV models, it is most convenient to partition the information in volatility innovations into that related to return news and that specific to volatility. This is implemented by decomposing the volatility innovation ξ_t into its projection on the return innovation ϵ_t and the projection error given by $E(\xi_t|\epsilon_t)$ and $v\,\eta_t$ respectively

$$\xi_t = E(\xi_t|\epsilon_t) + v\eta_t$$

where η_t has unit variance. So the volatility equation of the SV model becomes

$$\log(\sigma_{t+1}^2) = \gamma + \delta \log(\sigma_t^2) + E(\xi_t|\epsilon_t) + v\eta_t \tag{3.2}$$

The existing SV models differ in their treatment of the projection term $E(\xi_t|\epsilon_t)$. The BSV model assumes $E(\xi_t|\epsilon_t) = 0$. So the volatility process is simply given by

$$\log(\sigma_{t+1}^2) = \gamma + \delta \log(\sigma_t^2) + v\eta_t \tag{3.3}$$

and there is no correlation between the return and volatility innovations.

The SVL model in Harvey and Shephard (1996) and Omori, Chib, Shephard, and Nakajima (2007) amounts to assuming $E(\xi_t|\epsilon_t)$ as the linear least-squares projection of ξ_t on ϵ_t, that is, $E(\xi_t|\epsilon_t) = \text{var}(\epsilon_t)^{-1}\text{cov}(\xi_t, \epsilon_t)\ \epsilon_t$. Let $\rho \equiv \text{var}(\epsilon_t)^{-1}\text{cov}(\xi_t, \epsilon_t)$. The correlation coefficient between the return innovation ϵ_t and the log volatility innovation ξ_t is given by the constant ρ/σ_ξ. The SVL model can be written as

$$\log(\sigma_{t+1}^2) = \gamma + \delta \log(\sigma_t^2) + \rho\epsilon_t + v\eta_t \tag{3.4}$$

If $\rho < 0$, the return innovation ϵ_t and the volatility σ_{t+1}^2 are negatively correlated. Note that although according to Equation (3.4) negative and positive return shocks ϵ_t have equal effects on the log volatility $\log(\sigma_{t+1}^2)$, the volatility σ_{t+1}^2 will nevertheless have larger response to negative return shocks than to positive ones of the same magnitude. For example, suppose $\gamma + \delta \log(\sigma_t^2) + v\eta_t = 0$ in Equation (3.4). Let $\epsilon_t = \Delta > 0$ be a positive return shock. We will have $\log(\sigma_{t+1}^2) = \rho\Delta$ and hence $\sigma_{t+1}^2 = \exp(\rho\Delta)$. An

equally large negative return shock $\epsilon_t = -\Delta$ will result in $\log(\sigma_{t+1}^2) = -\rho\Delta$ and hence $\sigma_{t+1}^2 = \exp(-\rho\Delta)$. Since the coefficient ρ in Equation (3.4) is usually found to be negative in practice, we have $\rho\Delta < 0$ which implies $\exp(-\rho\Delta) > \exp(\rho\Delta)$. So the negative return shock $\epsilon_t = -\Delta$ will generate larger impact on the volatility σ_{t+1}^2 than the equally large positive return shock $\epsilon_t = \Delta$. The magnitude of the asymmetry, however, is rigidly determined by the logarithmic functional form.

The EGARCH model of Nelson (1991), which is given by

$$\log(\sigma_{t+1}^2) = \gamma + \delta \log(\sigma_t^2) + \psi\epsilon_t + \phi\left(|\epsilon_t| - E\left(|\epsilon_t|\right)\right) \tag{3.5}$$

can also be regarded as a member of the models described by Equation (3.2) in which $E(\xi_t|\epsilon_t) = \psi\epsilon_t + \phi\left(|\epsilon_t| - E\left(|\epsilon_t|\right)\right)$ and $v = 0$, that is, the projection of ξ_t on ϵ_t is a nonlinear function of ϵ_t and the projection error is degenerate. The EGARCH model entertains a richer structure of relations between the return innovation ϵ_t and the volatility innovation ξ_t than the SVL model since the covariance between ϵ_t and ξ_t can take different values depending on the sign of ϵ_t. But one limitation of the EGARCH model is that conditional on the sign of ϵ_t, ξ_t becomes a linear function of ϵ_t and hence the correlation between ϵ_t and ξ_t is either 1 or -1. This could be too restrictive in practice and is due to the fact that there is only a single source of shocks in the EGARCH model. As will be shown below, there are strong theoretical and empirical grounds to support the inclusion of volatility-specific shocks when studying the leverage effect of stock returns.

This chapter proposes a tilt stochastic volatility (TSV) model which extends the existing volatility models to include both nonlinear structure and volatility-specific innovations

$$\log(\sigma_{t+1}^2) = \gamma + \delta \log(\sigma_t^2) + \psi\epsilon_t + \phi|\epsilon_t| + v\eta_t \tag{3.6}$$

where the term $\phi|\epsilon_t|$ is referred to as the tilt effect and ϕ as the tilt coefficient. It can be obtained from Equation (3.2) by setting $E(\xi_t|\epsilon_t) = \psi\epsilon_t + \phi\left(|\epsilon_t| - E\left(|\epsilon_t|\right)\right)$ and adjusting the intercept term of the equation accordingly. The TSV model nests both the BSV

model, the SVL model and the EGARCH model as special cases: when $\psi = \phi = 0$, it corresponds to the BSV model; when $\phi = 0$, it becomes the SVL model and when $v = 0$, it reduces to the EGARCH model. The correlation coefficient between the return innovation ϵ_t and the log volatility innovation ξ_t is given by

$$\rho = \begin{cases} \frac{\psi+\phi}{\sqrt{(\psi+\phi)^2+v^2}}, & \text{when } \epsilon_t \geq 0; \\ \frac{\psi-\phi}{\sqrt{(\psi-\phi)^2+v^2}}, & \text{when } \epsilon_t < 0. \end{cases} \tag{3.7}$$

The difference between the BSV, the SVL and the TSV models in describing the leverage effect is illustrated by Figure (3.1), which relates the magnitude of changes in volatility to return innovations, i.e. plots $|\sigma_t^2(\epsilon_t) - \sigma_t^2(\epsilon_t = 0)|$ against ϵ_t. By adjusting the tilt coefficient ϕ, the TSV model can generate more flexible asymmetric responses of volatility to return shocks than the other two SV models.

The difference between the TSV model and the EGARCH model lies in that the TSV model allows for volatility-specific shocks which are driven by news unrelated to stock price. To see why it is important to include volatility-specific shocks, consider the leverage hypothesis of Black (1976). It argues that when the price of stock drops, financial leverage is increased, which makes the stock riskier and hence increases its volatility. The term $E(\xi_t|\epsilon_t)$ can be interpreted as capturing the effect put forward in this hypothesis. But as Christie (1982) and Schwert (1989) show, this effect explains only a small portion of the fluctuations of volatility over time and many other macro variables such as interest rates have important influence on stock volatility. Also, as shown in French and Roll (1986) and Schwert (1989), variables of trading activity such as the share trading volume growth are closely related to stock volatility. So it is necessary to introduce volatility-specific shocks in order to account for the magnitude of changes in volatility over time and to incorporate the effects of macro and trading variables.

3.3 Estimation

It is convenient to define $\lambda_t \equiv \log(\sigma_t^2)$. So the TSV model to be estimated can be rewritten as

$$\begin{aligned} y_t &= \exp(0.5\lambda_t)\epsilon_t \\ \lambda_{t+1} &= \gamma + \delta\lambda_t + \psi\epsilon_t + \phi|\epsilon_t| + v\eta_t \end{aligned} \tag{3.8}$$

Let $\theta \equiv [\gamma, \delta, \psi, \phi, v]'$ collect the model parameters. Denote $g(y_t|\lambda_t;\theta)$ and $p(\lambda_t|y_{t-1}, \lambda_{t-1};\theta)$ as the conditional densities of y_t and λ_t respectively, $t = 0, 1, ..., n$. For convenience of exposition, λ_0 is treated as fixed. Relaxing this assumption needs only minor modification of the discussions below. The resulting likelihood function is

$$\begin{aligned} L(y_n, ..., y_1;\theta) &= \int f(y_n, ..., y_1, \lambda_n, ..., \lambda_1;\theta)d\lambda_n \cdots d\lambda_1 \\ &= \int \prod_{t=1}^{n} f(y_t, \lambda_t|y_{t-1}, \lambda_{t-1}, ..., y_1, \lambda_1;\theta)d\lambda_n \cdots d\lambda_1 \\ &= \int \prod_{t=1}^{n} g(y_t|\lambda_t;\theta)p(\lambda_t|y_{t-1}, \lambda_{t-1};\theta)d\lambda_n \cdots d\lambda_1 \end{aligned} \tag{3.9}$$

3.3.1 The Efficient Importance Sampler

In principle, a natural Monte Carlo estimate of the likelihood function given in Equation (3.9) can be obtained by

$$\hat{L}_N(y_n, ..., y_1;\theta) = \frac{1}{S}\sum_{i=1}^{S}\prod_{t=1}^{n} g(y_t|\lambda_t^{(i)};\theta) \tag{3.10}$$

where $\left\{\lambda_t^{(i)}\right\}_{t=1}^{n}$ is a sample path of the latent variables $\{\lambda_t\}_{t=1}^{n}$ drawn from the sequence of conditional densities $p(\lambda_t|y_{t-1}, \lambda_{t-1};\theta)$. But in practice, this natural Monte Carlo estimator is highly inefficient. The required number of draws S to achieve a reasonably accurate estimate of $L(y_n, ..., y_1;\theta)$ has to increase dramatically with the sample size n and is generally prohibitively large. See Danielsson and Richard (1993) for an illustration of such inefficiency.

To improve efficiency, the natural sampler $p(\lambda_t|y_{t-1}, \lambda_{t-1}; \theta)$ can be replaced by an auxiliary density function, termed importance sampler, $m(\lambda_t|y_{t-1}, \lambda_{t-1}; \theta, a_t)$, where a_t is a vector of additional parameters introduced by the importance sampler. So the likelihood function can be rewritten as

$$L(y_n, ..., y_1; \theta) = \int \left[\prod_{t=1}^{n} \frac{g(y_t|\lambda_t; \theta) p(\lambda_t|y_{t-1}, \lambda_{t-1}; \theta)}{m(\lambda_t|y_{t-1}, \lambda_{t-1}; \theta, a_t)} m(\lambda_t|y_{t-1}, \lambda_{t-1}; \theta, a_t) \right] d\lambda_n \cdots d\lambda_1 \tag{3.11}$$

The resulting Monte Carlo estimate of the likelihood function is given by

$$\hat{L}_I(y_n, ..., y_1; \theta) = \frac{1}{S} \sum_{i=1}^{S} \prod_{t=1}^{n} \frac{g(y_t|\lambda_t^{(i)}; \theta) p(\lambda_t^{(i)}|y_{t-1}, \lambda_{t-1}^{(i)}; \theta)}{m(\lambda_t^{(i)}|y_{t-1}, \lambda_{t-1}^{(i)}; \theta, a_t)} \tag{3.12}$$

where $\left\{ \lambda_t^{(i)} \right\}_{t=1}^{n}$ is a sample path of the latent variables $\{\lambda_t\}_{t=1}^{n}$ drawn from the sequence of importance sampler $m(\lambda_t|y_{t-1}, \lambda_{t-1}; \theta, a_t)$. By judicially constructing the importance sampler to exploit the information of $\{\lambda_t\}_{t=1}^{n}$ implied by $\{y_t\}_{t=1}^{n}$, one can obtain accurate estimates of the likelihood function with a small number of Monte Carlo draws.

There have been a variety of innovative methods proposed to construct the importance sampler. See, for example, Tierney and Kadane (1986), Geweke (1989), Durbin and Koopmans (1997) and Owen and Zhou (2000) among many others. The method adopted in this chapter is the Efficient Importance Sampling (EIS) procedure of Richard and Zhang (2007), which is a highly generic Monte Carlo technique for evaluating very large dimensional integrals.

The intuition behind the EIS procedure is that, if we can find a sequence of density $m(\lambda_t|y_{t-1}, \lambda_{t-1}; \theta, a_t)$ proportional to the integrant $g(y_t|\lambda_t; \theta) p(\lambda_t|y_{t-1}, \lambda_{t-1}; \theta)$, the estimate $\hat{L}_I(y_n, ..., y_1; \theta)$ given in Equation (3.12) would be a constant and hence its sampling variance would be ideally zero. Since in general this is impossible, the EIS procedure aims at selecting parameters a_t to make $m(\lambda_t|y_{t-1}, \lambda_{t-1}; \theta, a_t)$ closely resemble $g(y_t|\lambda_t; \theta) p(\lambda_t|y_{t-1}, \lambda_{t-1}; \theta)$ up to a scaler so to minimize the sampling variance of

$\hat{L}_I(y_n, ..., y_1; \theta)$. This is achieved by choosing a_t to minimize the function

$$\sum_{i=1}^{S} \left[\log \left(g(y_t|\lambda_t^{(i)}; \theta) p(\lambda_t^{(i)}|y_{t-1}, \lambda_{t-1}^{(i)}; \theta) \right) - c_t - \log \left(m(\lambda_t^{(i)}|y_{t-1}, \lambda_{t-1}^{(i)}; \theta, a_t) \right) \right]^2$$

where $c_t \in a_t$ is a parameter capturing the scalar difference between the integrant and the importance sampler. In practice, the EIS parameters in a_t are determined by iterating the minimization several times. The latent variable $\left\{ \lambda_t^{(i)} \right\}_{t=1}^{n}$ is initially drawn from the natural sampler $p(\lambda_t|y_{t-1}, \lambda_{t-1}; \theta)$ and subsequently drawn from the importance sampler $m(\lambda_t|y_{t-1}, \lambda_{t-1}; \theta, a_t)$ with a_t estimated from the previous round. A modified method which greatly improves efficiency is to introduce the integrating constant

$$\chi(y_{t-1}, \lambda_{t-1}; \theta, a_t) \equiv \frac{\kappa(\lambda_t, y_{t-1}, \lambda_{t-1}; \theta, a_t)}{m(\lambda_t|y_{t-1}, \lambda_{t-1}; \theta, a_t)}$$

where $\kappa(\lambda_t, y_{t-1}, \lambda_{t-1}; \theta, a_t)$ is the kernel of the conditional density $m(\lambda_t|y_{t-1}, \lambda_{t-1}; \theta, a_t)$. The problem is reformulated as sequentially choosing a_t, from n to 1, to minimize the function

$$\sum_{i=1}^{S} \left[\log \left(g(y_t|\lambda_t^{(i)}; \theta) p(\lambda_t^{(i)}|y_{t-1}, \lambda_{t-1}^{(i)}; \theta) \chi(y_t, \lambda_t^{(i)}; \theta, \hat{a}_{t+1}) \right) - c_t - \log \left(\kappa(\lambda_t^{(i)}|y_{t-1}, \lambda_{t-1}^{(i)}; \theta, a_t) \right) \right]^2$$

where $\chi(y_n, \lambda_n; \theta, a_{n+1}) \equiv 1$.

Besides its high numerical accuracy, the EIS procedure has several attractive features. It is very flexible, well suited for large-dimensional integration problems and hence able to accommodate complicated model structures. Its implementation is relatively straightforward and the computation cost is comparatively low. See Richard and Zhang (2007) for a more detailed discussion of the EIS procedure.

3.3.2 Estimation of the Tilt SV Model

Assume the return innovation ϵ_t and the volatility-specific innovation η_t are normally distributed. So the conditional densities are given by

$$g(y_t|\lambda_t; \theta) \propto \exp \left(-\frac{\lambda_t + y_t^2 \exp(-\lambda_t)}{2} \right)$$

$$p(\lambda_t|y_{t-1},\lambda_{t-1};\theta) \propto \exp\left(-\frac{[\lambda_t-\gamma-\delta\lambda_{t-1}-\psi y_{t-1}\exp(-0.5\lambda_{t-1})-\phi|y_{t-1}\exp(-0.5\lambda_{t-1})|]^2}{2v^2}\right) \tag{3.13}$$

Let $a_t=(a_{1t},a_{2t},c_t)$ and $d(\lambda_t;a_t)\equiv\exp(a_{1t}\lambda_t+a_{2t}\lambda_t^2)$ be the kernel of $p(\lambda_t|y_{t-1},\lambda_{t-1};\theta)$. The kernel of the importance sampler $m(\lambda_t|y_{t-1},\lambda_{t-1};\theta,a_t)$ is constructed as $p(\lambda_t|y_{t-1},\lambda_{t-1};\theta)d(\lambda_t;a_t)$ and is given by

$$\kappa(\lambda_t,y_{t-1},\lambda_{t-1};\theta,a_t) \propto \exp\left(-\frac{1}{2}\left[\frac{h_t^2}{v^2}-2\left(\frac{h_t}{v^2}+a_{1t}\right)\lambda_t+\left(\frac{1}{v^2}-2a_{2t}\right)\lambda_t^2\right]\right) \tag{3.14}$$

where $h_t\equiv\gamma+\delta\lambda_{t-1}+\psi y_{t-1}\exp(-0.5\lambda_{t-1})+\phi|y_{t-1}\exp(-0.5\lambda_{t-1})|$.

The resulting importance sampler $m(\lambda_t|y_{t-1},\lambda_{t-1};\theta,a_t)$ is a normal density function with mean μ_t and variance ω_t^2 given by

$$\mu_t=\omega_t^2\left(\frac{h_t}{v^2}+a_{1t}\right),\quad \omega_t^2=\frac{v^2}{1-2v^2a_{2t}}$$

The integrating constant takes the form

$$\chi(y_{t-1},\lambda_{t-1};\theta,a_t) \propto \exp\left(\frac{\mu_t^2}{2\omega_t^2}-\frac{h_t^2}{2v^2}\right)$$

The Maximum Likelihood (ML)-ELS estimation of the tilt SV model includes the following steps.

1. Use the natural sampler $p(\lambda_t|y_{t-1},\lambda_{t-1};\theta)$ to draw independently S sample paths of the latent variable $\{\lambda_t^{(i)}\}_{t=1}^n, i=1,2,...,S$.

2. Compute the first-round estimates of the EIS parameters $\{\hat{a}_t\}_{t=1}^n$ by the EIS-Least Squares procedure

$$\hat{a}_t=\arg\max_{a_t}\sum_{i=1}^{S}$$
$$\left[\log\left(g(y_t|\lambda_t^{(i)};\theta)p(\lambda_t^{(i)}|y_{t-1},\lambda_{t-1}^{(i)};\theta)\chi(y_t,\lambda_t^{(i)};\theta,\hat{a}_{t+1})\right)-c_t-\log\left(\kappa(\lambda_t^{(i)},y_{t-1},\lambda_{t-1}^{(i)};\theta,a_t)\right)\right]^2$$
$$=\arg\max_{a_t}\sum_{i=1}^{S}$$
$$\left[-\frac{\lambda_t^{(i)}+y_t^2\exp(-\lambda_t^{(i)})}{2}+\frac{1}{2}\left(\frac{\left(\mu_{t+1}^{(i)}(\hat{a}_{t+1})\right)^2}{\omega_{t+1}^2(\hat{a}_{t+1})}-\frac{\left(h_{t+1}^{(i)}\right)^2}{v^2}\right)-c_t-a_{1t}\lambda_t^{(i)}-a_{2t}\left(\lambda_t^{(i)}\right)^2\right]^2$$

where $\chi(y_n, \lambda_n; \theta, a_{n+1}) \equiv 1$, that is, we regress

$$-\frac{\lambda_t^{(i)} + y_t^2 \exp(-\lambda_t^{(i)})}{2} + \frac{1}{2}\left(\frac{\left(\mu_{t+1}^{(i)}\left(\hat{a}_{t+1}\right)\right)^2}{\omega_{t+1}^2\left(\hat{a}_{t+1}\right)} - \frac{\left(h_{t+1}^{(i)}\right)^2}{v^2}\right)$$

on $\left[1 \ \lambda_t^{(i)} \ \left(\lambda_t^{(i)}\right)^2\right]$ to obtain estimates of the parameters $a_t = (c_t, a_{1t}, a_{2t})$.

3. Use the importance sampler with the first-round EIS parameters $m(\lambda_t|y_{t-1}, \lambda_{t-1}; \theta, \hat{a}_t)$ to draw independently S sample paths of the latent variable $\{\lambda_t^{(i)}\}_{t=1}^n, i = 1, 2, ..., S$.

4. Repeat the EIS-Least Squares procedure to obtain the second-round estimates of the EIS parameters $\{\hat{a}_t\}_{t=1}^n$.

5. Repeat the above steps q times to obtain the qth-round estimates of the EIS parameters and let them be the final EIS parameter estimates.

6. Use the importance sampler with the final estimates of the EIS parameters to draw independently S sample paths of the latent variable $\{\lambda_t^{(i)}\}_{t=1}^n, i = 1, 2, ..., S$.

7. Compute the log likelihood

$$\log \hat{L}_I(y_n, ..., y_1; \theta) = \log\left(\frac{1}{S}\sum_{i=1}^{S}\prod_{t=1}^{n-1}\frac{g(y_t|\lambda_t^{(i)}; \theta)p(\lambda_t^{(i)}|y_{t-1}, \lambda_{t-1}^{(i)}; \theta)}{m(\lambda_t^{(i)}|y_{t-1}, \lambda_{t-1}^{(i)}; \theta, \hat{a}_t)}\right)$$

8. Compute the single-round ML-EIS estimate

$$\hat{\theta} = \arg\ \max_{\theta} \log \hat{L}_I(y_n, ..., y_1; \theta)$$

9. Repeat the above steps R times. Denote the single-round ML-EIS estimate as $\hat{\theta}_r, r = 1, 2, ..., R$. The final ML-EIS estimate is $\hat{\theta} = \frac{1}{R}\sum_{r=1}^{R}\hat{\theta}_r$.

It is very important to note that the Common Random Numbers technique (CRN) should be used to draw the latent variables, that is, the latent variables $\{\lambda_t^{(i)}\}_{t=1}^n, i = 1, 2, ..., S$, from $m(\lambda_t|y_{t-1}, \lambda_{t-1}; \theta, \hat{a}_t)$ should be obtained from a *common* sequence of random numbers $\{\widetilde{u}_t^{(i)}\}_{t=1}^n, i = 1, 2, ..., S$, whose draw does not depend on θ. Otherwise

the log likelihood function $\log \hat{L}_I(y_n, ..., y_1; \theta)$ would exhibit serious discontinuity as the computer searches over different values of θ in optimization routines and the numerical maximization would be highly inefficient. In this chapter, the common random numbers $\{\widetilde{u}_t^{(i)}\}_{t=1}^n, i = 1, 2, ..., S$, is simply a $S \times n$ matrix of uncorrelated standard normal random numbers. The latent variables $\lambda_t^{(i)}$, are constructed as $\lambda_t^{(i)} = \mu_t + \omega_t \widetilde{u}_t^{(i)}$.

Once the ML-EIS estimates are obtained, two distinct covariance matrices of the estimates are to be computed: the numerical covariance matrix and the statistical covariance matrix. The numerical covariance matrix evaluates the numerical accuracy of the estimates. Let θ^* be the maximizer of the (infeasible) analytical log likelihood function $\log L(y_n, ..., y_1; \theta)$. Since the ML-EIS estimate $\hat{\theta}$ is the maximizer of $\log \hat{L}_I(y_n, ..., y_1; \theta)$ which is a numerical approximation of the log likelihood function, we need a measure of how accurate $\hat{\theta}$ is as an approximation of θ^*. This is given by the standard errors computed from the numerical covariance matrix, whose formula is $\frac{1}{R}\sum_{r=1}^R \hat{\theta}_r \hat{\theta}_r' - \hat{\theta}\hat{\theta}'$.

On the other hand, even if $\hat{\theta}$ is a perfect approximation of θ^*, i.e. $\hat{\theta} = \theta^*$, there is still statistical randomness since θ^* is just an estimate of the true parameters θ and hence is subject to statistical errors. The statistical covariance matrix is used to evaluate this statistical randomness. All conventional statistical inferences are to be conducted using the statistical covariance matrix. It is computed as follows. Treat $\hat{\theta}$ as the true parameter and draw independently L sets of $\{y_t^{(i)}, \lambda_t^{(i)}\}_{t=1}^n, i = 1, 2, ..., L$, from model (3.8). For each set of $\{y_t^{(i)}, \lambda_t^{(i)}\}_{t=1}^n$, we retain only $\{y_t^{(i)}\}_{t=1}^n$ and use the EIS procedure to compute numerically the log likelihood $\log \hat{L}_I(y_n, ..., y_1; \theta)$. Denote the resulting estimate as $\hat{\theta}^{(i)} = \arg \max_\theta \log \hat{L}_I(y_n^{(i)}, ..., y_1^{(i)}; \theta)$. The statistical covariance matrix is given by $\frac{1}{L}\sum_{i=1}^L \left(\hat{\theta}^{(i)}\right)\left(\hat{\theta}^{(i)}\right)' - \left(\frac{1}{L}\sum_{i=1}^L \hat{\theta}^{(i)}\right)\left(\frac{1}{L}\sum_{i=1}^L \hat{\theta}^{(i)}\right)'$.

It should be noted that the EIS practice of computing both numerical and statistical covariance matrices contrasts with the conventional treatment of simulation estimators, e.g. Gourieroux and Monfort (1996), in which the sampling property of simulation estimators is derived from the joint distribution of the observations and latent variables.

See Richard and Zhang (2007) for an argument in favor of the EIS procedure in terms of simulation efficiency.

3.3.3 Volatility Estimates

Once the parameters of the TSV model are estimated, the smoothed volatility can be computed as

$$\hat{\sigma}^2_{t,S} = \hat{E}\left(\exp(\lambda_t)|y_1, ..., y_n\right) = \frac{1}{S}\sum_{i=1}^{S} w_{i,S} \exp\left(\lambda_t^{(i)}\right)$$

where

$$w_{i,S} \equiv \frac{\prod_{\tau=1}^{n} \frac{g(y_\tau|\lambda_\tau^{(i)};\hat{\theta})p(\lambda_\tau^{(i)}|y_{\tau-1},\lambda_{\tau-1}^{(i)};\hat{\theta})}{m(\lambda_\tau^{(i)}|y_{\tau-1},\lambda_{\tau-1}^{(i)};\hat{\theta},\hat{a}_\tau)}}{\sum_{j=1}^{S}\prod_{\tau=1}^{n} \frac{g(y_\tau|\lambda_\tau^{(j)};\hat{\theta})p(\lambda_\tau^{(j)}|y_{\tau-1},\lambda_{\tau-1}^{(j)};\hat{\theta})}{m(\lambda_\tau^{(j)}|y_{\tau-1},\lambda_{\tau-1}^{(j)};\hat{\theta},\hat{a}_\tau)}}$$

and $\lambda_\tau^{(i)}$ is the τth element of the ith sample path of log volatilities drawn from the EIS importance sampler.

In principle, the filtered volatility can be computed in a similar manner

$$\hat{\sigma}^2_{t,F} = \hat{E}\left(\exp(\lambda_t)|y_1, ..., y_t\right) = \frac{1}{S}\sum_{i=1}^{S} w_{i,F} \exp\left(\lambda_t^{(i)}\right)$$

where

$$w_{i,F} \equiv \frac{\prod_{\tau=1}^{t} \frac{g(y_\tau|\lambda_\tau^{(i)};\hat{\theta})p(\lambda_\tau^{(i)}|y_{\tau-1},\lambda_{\tau-1}^{(i)};\hat{\theta})}{m(\lambda_\tau^{(i)}|y_{\tau-1},\lambda_{\tau-1}^{(i)};\hat{\theta},\hat{a}_\tau)}}{\sum_{j=1}^{S}\prod_{\tau=1}^{t} \frac{g(y_\tau|\lambda_\tau^{(j)};\hat{\theta})p(\lambda_\tau^{(j)}|y_{\tau-1},\lambda_{\tau-1}^{(j)};\hat{\theta})}{m(\lambda_\tau^{(j)}|y_{\tau-1},\lambda_{\tau-1}^{(j)};\hat{\theta},\hat{a}_\tau)}}$$

and $\lambda_\tau^{(i)}$ is the τth element of the ith sample path of log volatilities drawn from the EIS importance sampler. In this case, the EIS parameters $\{\hat{a}_\tau\}_{\tau=1}^{t}$ need to be recomputed for each $t = 1, 2, ..., n$. Since the algorithm determining the EIS parameters $\{\hat{a}_\tau\}_{\tau=1}^{t}$ runs backward from t to 1, it can be computationally intensive when the data size n is large. This chapter only reports the smoothed estimates of volatility for SV models. A practical implication is that the results in this chapter should be more appropriately interpreted as *ex post* description of data than *ex ante* forecast in portfolio selection context.

3.4 Simulation Study

In this section, I use Monte Carlo (MC) simulations to investigate the small sample properties of the ML-EIS estimator for both the SVL and TSV models.

The simulation data is generated according to Equation (3.8). In particular, two sets of parameters, which I consider empirically plausible, are selected for simulation (Table 3.1). Under the first set of parameters, the SVL model represents the "true" model while the TSV model is misspecified and unnecessarily estimates a tilt coefficient. The reverse is true for the second set of parameters. The aim of the simulation study is to see and compare how the ML-EIS estimator behaves under a "true" model and a misspecified model so to provide guidance for empirical application.

The number of Monte Carlo repetitions is 100, which is relatively small by the standards of usual simulation studies. But considering the high accuracy of the ML-EIS estimator, we think that such a number of Monte Carlo repetitions is able to adequately describe its finite sample performance. In each MC repetition, a sample of 1500 observations is simulated and is estimated for both the SVL model and the TSV model according to the ML-EIS procedure described in the preceding section with $q = 3$ and $S = 10$. For each simulated sample, the SVL model and the TSV model are estimated with the same set of CRN so their comparison will not be affected by the difference in their respective random number draws. It should also be noted that the sample size of 1500 is small for financial time series analysis and is chosen to test the precision of the ML-EIS estimator.

Two measures of accuracy for parameter estimates are used in the simulation study: the mean absolute error (MAE) and the root mean squared error (RMSE) defined as

$$MAE = \frac{1}{100}\sum_{i=1}^{100}|\hat{\theta}_i - \theta|$$

$$RMSE = \sqrt{\frac{1}{100}\sum_{i=1}^{100}\left(\hat{\theta}_i - \theta\right)^2}$$

respectively, where θ is the true parameter and $\hat{\theta}_i$ is the ML-EIS estimate in the ith Monte Carlo repetition.

I also consider a measure of accuracy for volatility estimates: the mean of RMSE of volatility (MRMSE) defined as

$$MRMSE = \frac{1}{100}\sum_{i=1}^{100} RMSE_i$$

where $RMSE_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\sigma}_{t,S}^{2(i)} - \sigma_t^{2(i)}\right)^2}$, $\sigma_t^{2(i)}$ is the "true" volatility simulated in the ith MC repetition, $\hat{\sigma}_{t,S}^{2(i)}$ is the smoothed estimate of $\sigma_t^{2(i)}$ and T is the sample size.

Table 3.2 and 3.3 report the MAE, RMSE and MRMSE for estimates under the two set of parameters. For a sample size as small as 1500, all of the parameters and volatility are estimated with reasonable accuracy. The RMSE of parameter estimates for the SVL model appear considerably smaller than those reported for the Quasi Maximum Estimator in the simulation study of Harvey and Shephard (1996).

The simulation results highlight the importance of model specification when estimating SV models. Under either set of parameters, the "true" model performs fairly better than the misspecified model for estimating all parameters and volatility. In particular, when the tilt effect is present, the TSV model produces markedly more accurate estimates for both parameters and volatility than the SVL model which fails to incorporate the tilt effect.

3.5 Empirical Application

3.5.1 Parameter Estimates

The data analyzed in this chapter is the Center for Research in Security Prices (CRSP) daily returns on a value-weighted U.S. market index from July 4, 1962 through December 31, 1987, which contains a total of 6409 observations. The summary statistics of the data series are provided in Table 3.4. The same dataset has been used in Nelson (1991),

Harvey and Shephard (1996) and Jacquier, Polson, and Rossi (2004). The purpose of using this dataset is to facilitate comparison with previous studies. Table 3.5 presents the estimation results for this series using the BSV, the SVL, the TSV and the EGARCH models.

Geweke (1989) argued that the convergence of importance sampling estimators will be slow and unstable if the variance of the importance weights does not exist. See also Robert and Casella (2004) and Koopman, Shephard, and Creal (2009). To check this condition, Figure 3.2 provides plots and recursive variance of the importance weights for the TSV model. These diagnostic graphs indicate that the efficient importance sampling is reasonably well behaved. In particular, there is no evidence that the variance of the importance weights does not exist.

The persistence parameter δ is estimated to be between 0.98 and 0.99 for all of the 4 models, consistent with the findings in Nelson (1991), Harvey and Shephard (1996) and Jacquier, Polson, and Rossi (2004). For the SV models, adding more features of leverage effect seems to slightly lower the value of δ.

The estimated leverage effects differ markedly across models. Figure 3.3 plots the projection curves based on the estimates in Table 3.5. For the SVL model, the implied correlation coefficient between return and volatility innovations ρ is estimated to be -0.42, which is slightly lower in absolute value than the estimate of -0.48 in Jacquier, Polson, and Rossi (2004) whereas substantially lower than the value of -0.66 reported in Harvey and Shephard (1996). But as seen from estimates of the TSV model, the tilt coefficient is significantly positive with a t statistic of 2.03. The implied correlation coefficients between return innovations and volatility innovations, as given in Equation (3.7), are -0.40 when the return innovation $\epsilon_t \geq 0$ and -0.59 when the return innovation $\epsilon_t < 0$. So there is considerable asymmetry missing in estimates of the SVL model. As shown in the next subsection, failure to take into account this tilt effect markedly deteriorates estimates of volatility.

Comparing with the EGARCH model, the TSV model has a very low estimate of 0.0185 of the tilt coefficient ϕ, which is an order of magnitude smaller than the EGARCH estimate of 0.1666. This arises because the EGARCH model confounds the effect of return shocks with that of volatility-specific shocks which it fails to take into account. Intuitively, in order to "catch up" with large fluctuations in volatility, the EGARCH model has to resort to a large value of ϕ to amplify the effect of return shocks while the TSV model, by allowing for volatility-specific shocks, needs only a small ϕ to fit the magnitude of volatility fluctuations. The implication is that volatility estimates in the EGARCH model would overreact to negative return shocks, which makes the volatility estimates very noisy and hence deteriorates their accuracy.

In summary, the most significant difference across the models lies in the estimates of leverage effect. In particular, the SVL model tends to underestimate the magnitude of the asymmetric correlation between return and volatility innovations while the EGARCH model seems overestimating this asymmetry. This difference, as shown in the next subsection, has important implications for their estimates of volatility.

3.5.2 Smoothed Volatility Estimates

In this section, we turn to examine the accuracy of the smoothed volatility estimates produced by the basic SV model, the SVL model, the TSV model and the EGARCH model.

Table 3.6 shows the sample mean of squared returns and volatility estimates by the 4 models. The statistic we consider is the sample mean of the difference between squared returns and volatility estimates $\frac{1}{n}\sum_{t=1}^{n}(y_t^2 - \hat{\sigma}_t^2)$. According to the law of large numbers, any unbiased estimates of volatility should converge to the mean of squared return. So this statistic should converge to zero for any unbiased volatility estimates. The asymptotic variance of this statistic is computed by the Newey-West method (Newey and West (1987)) to account for possible serial correlations in $y_t^2 - \hat{\sigma}_t^2$. The third row of Table

3.6 presents the values of the statistic with the Newey-West standard errors given in parentheses.

Among the three SV models, estimates by the TSV model turn out to be the best in terms of matching the sample mean of squared returns. This superiority is further confirmed by the statistics presented in Table 3.6. Although all of the volatility estimates pass the unbiasedness test, estimates by the BSV model and the SVL model show greater bias than those by the TSV model, which confirms the findings in the simulation study that ignoring the tilt effect makes the volatility estimates less accurate. It should also be noted that estimates by the EGARCH model are more accurate on average than the SV models. This is likely to be the result of its relatively greater parsimony.

The other measure of accuracy considered is the mean squared error (MSE) between squared returns and volatility estimates $\frac{1}{n}\sum_{t=1}^{n}\left((y_t^2-\hat{\sigma}_t^2)^2\right)$. To see the usefulness of this measure, let $u_t \equiv y_t^2-\sigma_t^2$ be the error between squared return y_t^2 and the true volatility σ_t^2, $v_t \equiv \sigma_t^2-\hat{\sigma}_t^2$ be the error between the true volatility σ_t^2 and its parametric model estimate $\hat{\sigma}_t^2$. Note that $E(v_t)$ is not necessarily zero since $\hat{\sigma}_t^2$ can be a biased estimate of σ_t^2. Assume the correlation between u_t and v_t is negligible, it can be easily shown

$$E\left(\left(y_t^2-\hat{\sigma}_t^2\right)^2\right)=E\left(u_t^2\right)+E\left(v_t^2\right)$$

So $E\left((y_t^2-\hat{\sigma}_t^2)^2\right)$ can be a useful measure for evaluating the ability of volatility estimates $\hat{\sigma}_t^2$ to track the variations in volatility σ_t^2. In finite sample, $E\left((y_t^2-\hat{\sigma}_t^2)^2\right)$ becomes $\frac{1}{n}\sum_{t=1}^{n}\left((y_t^2-\hat{\sigma}_t^2)^2\right)$, which is the MSE used in this chapter.

Under standard regularity conditions, the MSE differential between volatility estimates from model i and j, MSE_i-MSE_j, follows a normal distribution asymptotically. The asymptotic variance of the MSE differential can be computed by the Newey-West procedure to account for possible serial correlations. By assessing if the MSE differential is significantly smaller or greater than zero, a test can be performed to compare the accuracy of volatility estimates by two models.

Table 3.7 presents results of these measures. Since starting from October 1987, squared return becomes unusually high relative to previous periods, we also consider these measures for two subsamples. The first subsample is from July 4, 1962 to September 30, 1987 and the second is from October 1, 1987 to December 31, 1987. Numbers in parentheses are the Newey-West standard errors of MSE differentials.

As shown in the table, the SV models perform uniformly better than the EGARCH model on these measures. For example, the MSE of the EGARCH volatility estimates for the sample from July 4, 1962 to September 30, 1987 is 1.396, which is 11% larger than that of the TSV volatility estimates. Within the SV family models, the best again is the TSV model in terms of the smallest MSE. This clearly shows the usefulness of introducing the tilt effect into the SV model.

The MSE differentials are insignificantly negative for the full sample. This is due to fact that the large standard errors associated with the MSE differentials for the second subsample from Oct. 1, 1987 to Dec. 31, 1987 substantially bias the standard errors for the full sample MSE differentials upward. When excluding this period, the MSE differentials become highly significant, indicating that volatility estimates by the TSV model have significantly smaller MSE than those by the SVL model and the EGARCH model.

The finding that the EGARCH volatility estimates track the variations in squared return poorly seems at first glance puzzling since from Table 3.6, we see that they are the most accurate estimates of squared return on average. This can be explained by the fact that, by confounding the volatility-specific shocks with return shocks, the EGARCH model overestimates the tilt coefficient. When there is a negative return shock, the EGARCH model would inflate its effect and produce large volatility estimate. Since volatility estimates are highly persistent in the EGARCH model, this inflated effect would be transmitted to future estimates of volatility and die out very slowly. But abnormally large squared returns in the actual data, on the other hand, turn out to revert to their

normal levels quickly, leaving the EGARCH volatility estimates staying overly large for a long period. The TSV estimates of volatility, though equally persistent, do not rise as much as the EGARCH estimates following a negative return shock and hence are able to track the squared returns more closely for subsequent periods.

This argument is illustrated by Figure 3.4. The figure plots the squared return, the EGARCH estimates of volatility and the TSV estimates of volatility for the period from October 16, 1987 to December 31, 1987. The day October 19, 1987 (the business day following October 16,1987) witnesses the largest negative return, -17% in our data, in the post-war history of U.S. stock market. As seen in the figure, squared return rises dramatically on October 19 but jumps back and forth and returns to its normal level very quickly. The volatility estimates by either model increase, though much less in magnitude than squared return, on October 20 and then decay slowly, staying overly high for a very prolonged periods. The EGARCH estimates of volatility rise more on October 20 and hence overestimate the subsequent volatility more seriously than the TSV model. This clearly demonstrates why the EGARCH estimates track squared return poorly. Volatility estimates by the TSV model are less sensitive to the abnormally large squared returns and hence trace squared returns more closely on average than those by the EGARCH models .

3.6 Conclusion

In this article, a new tilt stochastic volatility (TSV) model is proposed that generalizes the existing volatility models by studying the leverage effect in a unified and flexible framework. The TSV model directly estimates the magnitude of the asymmetric correlation between return and volatility innovations and incorporates volatility-specific shocks. The Efficient Importance Sampling (EIS) procedure of Richard and Zhang (2007) is adapted to estimate the SV models. Using simulated data, I find that the Maximum

Likelihood (ML)-EIS estimator produces accurate estimates for both parameters and latent volatilities. The empirical application studies the CRSP daily returns from July 4, 1962 to December 31, 1987. I find strong evidence for the extensions of the TSV model. The estimated leverage effect of the TSV model is markedly different from those of the SVL model and the EGARCH model. The SVL model underestimates the magnitude of the asymmetric relation between return and volatility innovations. The EGARCH model, on the other hand, overestimates the reaction of volatility to negative return innovations. Incorporating the extensions in the TSV model improves the accuracy of volatility estimates over the basic SV model and the SVL model. I also find that the volatility estimates of the TSV model track the variation in volatility more closely than the EGARCH estimates.

Table 3.1: Parameter Values for Simulation Study

	γ	δ	ψ	ϕ	v^2
Set 1	-0.1	0.95	-0.2	0	0.05
Set 2	-0.1	0.95	-0.2	0.1	0.05

Note: The model is $y_t = \exp(0.5\lambda_t)\epsilon_t, \quad \epsilon_t \sim N(0,1), \quad \lambda_{t+1} = \gamma + \delta\lambda_t + \psi\epsilon_t + \phi|\epsilon_t| + v\eta_t, \quad \eta_t \sim N(0,1)$. Under the 1st set of parameters, the correlation between the return innovation and the log volatility innovation is -0.66. Under the 2nd set of parameters, the correlation is -0.41 when the return innovation $\epsilon_t \geq 0$ and -0.80 when $\epsilon_t < 0$.

Table 3.2: Simulation Results: The First Set of Parameters

Panel A: Parameter Estimates					
	True value	TSV Estimates		SVL Estimates	
		MAE	RMSE	MAE	RMSE
γ	-0.1	0.0422	0.0493	0.0285	0.0343
δ	0.95	0.0127	0.0158	0.0097	0.0127
ψ	-0.2	0.0590	0.0601	0.0530	0.0555
ϕ	0	0.0179	0.0220		
v^2	0.05	0.0239	0.0291	0.0200	0.0234
Panel B: Volatility Estimates					
		TSV Estimates		SVL Estimates	
MRMSE		0.1391		0.1387	

Note: Estimates are based on 100 samples of 1500 observations simulated from Equation (3.8). The column "MAE" reports the mean absolute error of parameter estimates. The column "RMSE" reports the root mean squared error of parameter estimates. The row "MRMSE" reports the mean of root mean squared error of volatility estimates .

Table 3.3: Simulation Results: The Second Set of Parameters

Panel A: Parameter Estimates

	True value	TSV Estimates		SVL Estimates	
		MAE	RMSE	MAE	RMSE
γ	-0.1	0.0509	0.0576	0.0611	0.0649
δ	0.95	0.0104	0.0136	0.0105	0.0140
ψ	-0.2	0.0603	0.0687	0.0721	0.0798
ϕ	0.1	0.0826	0.0856		
v^2	0.05	0.0418	0.0443	0.0537	0.0562

Panel B: Volatility Estimates

	TSV Estimates	SVL Estimates
MRMSE	0.8703	0.9002

Note: Estimates are based on 100 samples of 1500 observations simulated from Equation (3.8). The column "MAE" reports the mean absolute error of parameter estimates. The column "RMSE" reports the root mean squared error of parameter estimates. The row "MRMSE" reports the mean of root mean squared error of volatility estimates .

Table 3.4: Summary Statistics of Daily CRSP Returns

Mean	Std. Deviation	Skewness	Kurtosis	Min.	Max.
0.042	0.824	-1.352	37.915	-17.135	8.662

This table reports the summary statistics of the daily CRSP returns from July 4, 1962 through December 31, 1987.

Table 3.5: Parameter Estimates for CRSP Daily Returns

	Basic SV	SVL	TSV	EGARCH
γ	-0.0133	-0.0144	-0.0267	-0.1390
	(0.0050)	(0.0063)	(0.0082)	(0.0204)
	[0.0013]	[0.0022]	[0.0016]	
δ	0.9853	0.9841	0.9830	0.9848
	(0.0012)	(0.0020)	(0.0015)	(0.0026)
	[0.0010]	[0.0007]	[0.0001]	
ψ		-0.0567	-0.0759	-0.0721
		(0.0046)	(0.0083)	(0.0106)
		[0.0025]	[0.0002]	
ϕ			0.0185	0.1666
			(0.0091)	(0.0261)
			[0.0020]	
v^2	0.0199	0.0152	0.0170	
	(0.0038)	(0.0057)	(0.0080)	
	[0.0006]	[0.0017]	[0.0001]	
log likelihood	-6697.93	-6627.08	-6614.37	-6687.63

Note: The CRSP daily returns are from July 4, 1962 to December 31, 1987 with a total of 6409 observations. Numbers in parentheses are statistical standard errors and numbers in square brackets are numerical standard errors of the ML-EIS estimates. The ML-EIS estimation of stochastic volatility models are based on 3 EIS iterations and 10 Monte Carlo repetitions. The EGARCH model is estimated via the maximum likelihood method assuming normally distributed return innovations. The log likelihoods of the SV models are the average across the Monte Carlo repetitions.

Table 3.6: Sample Mean of Squared Returns and Volatility Estimates

	y_t^2	$\hat{\sigma}^2_{t,BSV}$	$\hat{\sigma}^2_{t,SVL}$	$\hat{\sigma}^2_{t,TSV}$	$\hat{\sigma}^2_{t,EGARCH}$
Sample Mean	0.6786	0.6000	0.5976	0.6260	0.6646
$\frac{1}{n}\sum_{t=1}^{n}\left(y_t^2-\hat{\sigma}^2_{t,i}\right)$		-0.079	-0.081	-0.053	-0.014
		(0.061)	(0.060)	(0.054)	(0.053)

Note: y_t^2 is the squared return, $\hat{\sigma}^2_{t,i}$ is the volatility estimate produced by model i, i =BSV, SVL, TSV and EGARCH. Numbers in the parentheses are the Newey-West standard errors.

Table 3.7: Comparing Volatility Estimates

Panel A: MSE of Volatility Estimates				
Sample from July 4, 1962 to December 31, 1987				
	$\hat{\sigma}^2_{t,BSV}$	$\hat{\sigma}^2_{t,SVL}$	$\hat{\sigma}^2_{t,TSV}$	$\hat{\sigma}^2_{t,EGARCH}$
MSE	15.551	15.381	14.979	15.995
Sample from July 4, 1962 to September 30, 1987				
	$\hat{\sigma}^2_{t,BSV}$	$\hat{\sigma}^2_{t,SVL}$	$\hat{\sigma}^2_{t,TSV}$	$\hat{\sigma}^2_{t,EGARCH}$
MSE	1.274	1.295	1.262	1.396
Sample from October 1, 1987 to December 31, 1987				
	$\hat{\sigma}^2_{t,BSV}$	$\hat{\sigma}^2_{t,SVL}$	$\hat{\sigma}^2_{t,TSV}$	$\hat{\sigma}^2_{t,EGARCH}$
MSE	1431.033	1411.881	1374.926	1463.292
Panel B: MSE Differentials of Volatility Estimates				
Sample from July 4, 1962 to December 31, 1987				
	$MSE_{TSV} - MSE_{SVL}$		$MSE_{TSV} - MSE_{EGARCH}$	
MSE Differential	-0.402		-1.016	
	(0.344)		(0.775)	
Sample from July 4, 1962 to September 30, 1987				
	$MSE_{TSV} - MSE_{SVL}$		$MSE_{TSV} - MSE_{EGARCH}$	
MSE Differential	-0.032		-0.134	
	(0.014)		(0.018)	
Sample from October 1, 1987 to December 31, 1987				
	$MSE_{TSV} - MSE_{SVL}$		$MSE_{TSV} - MSE_{EGARCH}$	
MSE Differential	-36.989		-88.416	
	(28.212)		(74.450)	

Note: The rows "MSE" report the mean squared error between squared returns and volatility estimates. The rows "MSE Differential" report the difference between two MSEs. $\hat{\sigma}^2_{t,i}$ is the volatility estimate and MSE_i is the MSE by model i, i =BSV, SVL, TSV and EGARCH. Numbers in parentheses are the Newey-West standard errors of MSE differentials.

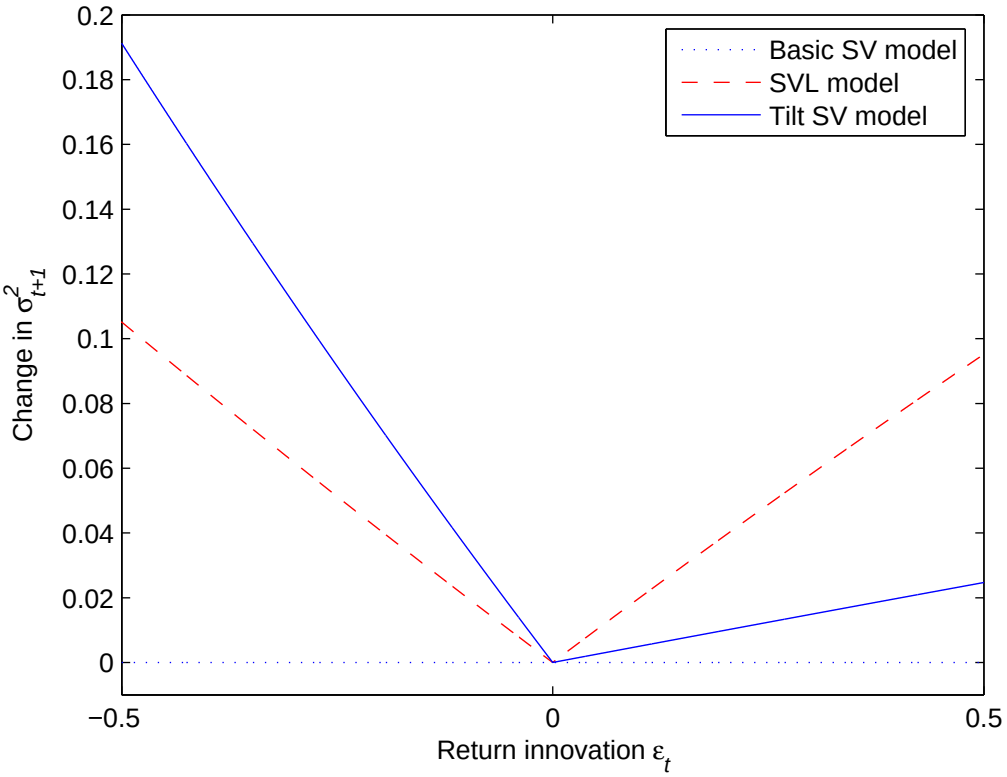

Figure 3.1: The Leverage Effect

Note: The plot illustrates the difference between the basic SV, the SVL and the tilt SV models in modeling the leverage effect. It is assumed that $\psi = -0.2$, $\phi = 0.15$ and $\sigma^2_{t+1} = 1$ when $\epsilon_t = 0$. The change in σ^2_{t+1} is defined as $|\sigma^2_{t+1} - 1|$.

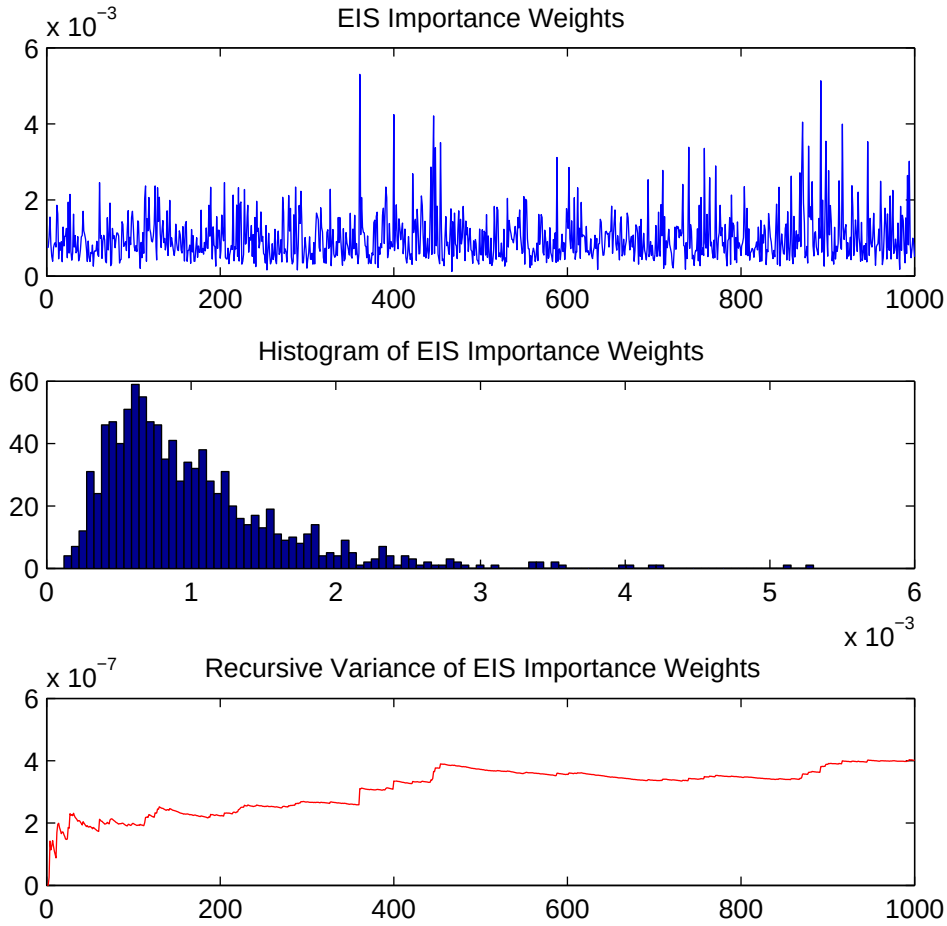

Figure 3.2: Importance Weights of The TSV Model

Note: The figures are based on 1000 sets of EIS importance weights simulated from the TSV model with parameters fixed at the EIS estimates.

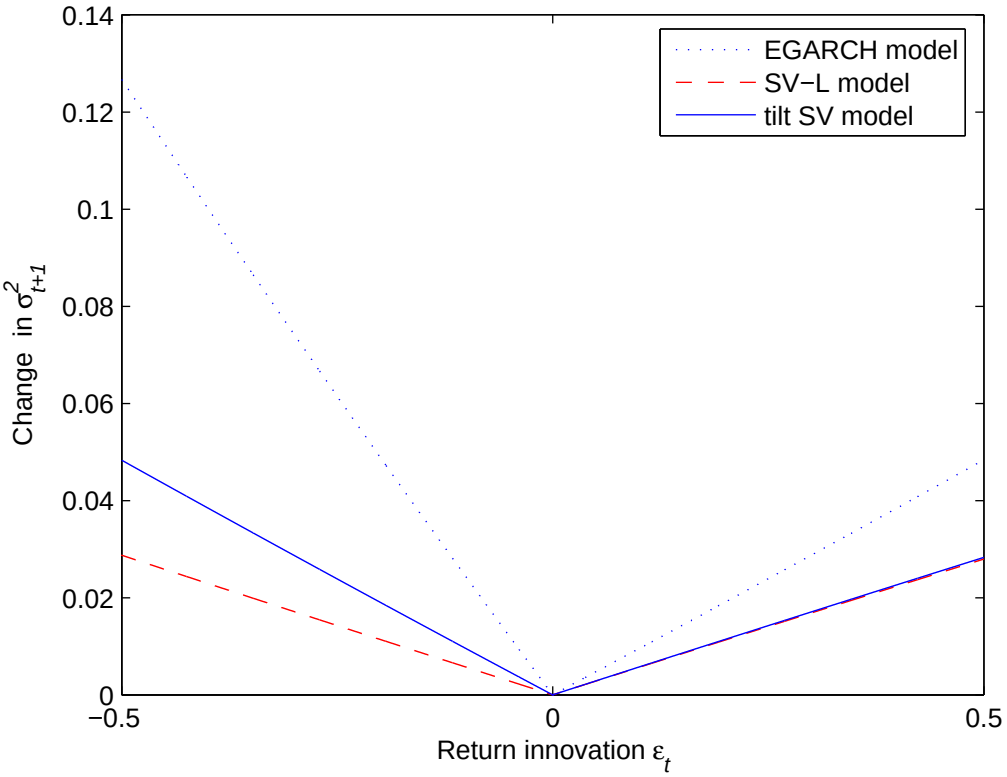

Figure 3.3: The Estimated Leverage Effect

Note: The plot shows the estimated leverage effects of the SVL model, the TSV model and the EGARCH model It is assumed that $\sigma^2_{t+1} = 1$ when $\epsilon_t = 0$. The change in σ^2_{t+1} is defined as $|\sigma^2_{t+1} - 1|$.

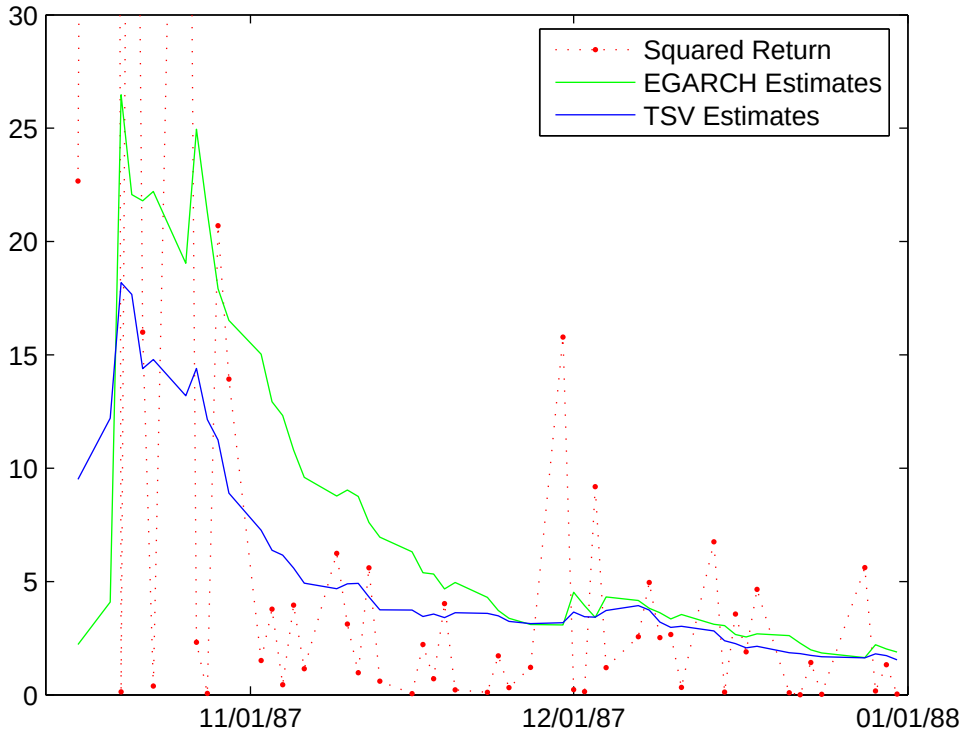

Figure 3.4: Squared Returns and Volatility Estimates

Note: The figure plots squared returns and volatility estimates by the EGARCH model and the TSV model for the period from October 16, 1987 to December 31, 1987. The squared returns above 0.003 are truncated to illustrate data more clearly.

Appendix A

Appendix to Chapter 1

A.1 Gibbs Sampler for the Structural Break VAR Model

For a given number of in-sample regimes K, let $\Theta = (\phi_1, ..., \phi_K, \Sigma_1, ..., \Sigma_K, \pi_1, ..., \pi_{K-1})$, $\Theta_0 = (b_0, B_0, \Omega_0, v_0, \alpha_0, \beta_0)$, $S = (s_1, s_2, ..., s_T)$ and $Y_T = (y_1, ..., y_T)$. The posterior distribution of interest is $p(S, \Theta, \Theta_0|Y_T)$. A Gibbs sampler is used to sample from this posterior distribution. Detailed discussions on Markov chain Monte Carlo methods of which Gibbs sampling is a special case can be found in Chib (2001), Koop (2003) and Geweke (2005).

1. Sample S from $p(S|Y_T, \Theta, \Theta_0)$.

Sampling the latent states S follows Chib (1998). The steps are as follows:

(1). Compute the filtering density $p(s_t|Y_t, \Theta)$ for $t = 1, 2, ..., T$ by the Hamilton filter (Hamilton (1989)). This involves repeatedly applying a prediction and a filtering step.

$$p(s_t = k|Y_{t-1}, \Theta) = p(s_{t-1} = k|Y_{t-1}, \Theta)\pi_k + p(s_{t-1} = k-1|Y_{t-1}, \Theta)(1 - \pi_{k-1})$$

$$p(s_t = k|Y_t, \Theta) \propto p(s_t = k|Y_{t-1}, \Theta)p(y_t|Y_{t-1}, s_t = k, \Theta)$$

for $k = 1, 2, ..., K$, where the data density $p(y_t|Y_{t-1}, s_t = k, \Theta) = N(y_t|\mu_k + \Phi_k y_{t-1}, \Sigma_k)$.

The filter starts at $p(s_t = 1|Y_0, \Theta) = 1$.

(2). Set $s_T = K$.

(3). Given $s_{t+1} = k$, sample s_t as

$$s_t = \begin{cases} \text{k}, & \text{with probability } c_t; \\ \text{k-1}, & \text{with probability } 1 - c_t. \end{cases}$$

where $c_t \propto p(s_t = k|Y_t, \Theta)p(s_{t+1} = k|s_t = k, \Theta)$, $t = T-1, T-2, ..., 2$. Note $s_1 = 1$.

2. Sample π_k from $p(\pi_k|Y_T, \Theta_{\_\pi_k}, \Theta_0, S)$ for $k = 1, ..., K-1$, where $\Theta_{\_\pi_k}$ is the subset of Θ excluding the parameter π_k.

$$\pi_k \sim Beta(\alpha_0 + n_{kk}, \beta_0 + 1)$$

where n_{kk} is the number of one-step transition from state k to state k in the sequence S.

3. Sample ϕ_k from $p(\phi_k|Y_T, \Theta_{\_\phi_k}, \Theta_0, S)$ and Σ_k from $p(\Sigma_k|Y_T, \Theta_{\_\Sigma_k}, \Theta_0, S)$, $k = 1, 2, ..., K$, where $\Theta_{\_\phi_k}$ and $\Theta_{\_\Sigma_k}$ are the subsets of Θ excluding the parameter ϕ_k and Σ_k respectively.

The conditional posterior densities of ϕ_k and Σ_k depend only on the data in regime k. Therefore, let $\{y_t : t = n_{k-1} + 1, n_{k-1} + 2, ..., n_k\}$ be the data in regime k where n_k is the last observation in regime k and $n_0 = 0$. Let $\widehat{Y}_k$ be a $(n_k - n_{k-1}) \times 3$ matrix of data with each row as an observation y_t, $\widehat{y}_k = vec(\widehat{Y}_k)$ be a vector stacking the columns of $\widehat{Y}_k$. Let $\widehat{x}_k$ be a $(n_k - n_{k-1}) \times 3$ matrix stacking observations $y_{n_{k-1}}, y_{n_{k-1}+1}, ..., y_{n_k-1}$, $\widehat{X}_k$ be a $(n_k - n_{k-1}) \times 4$ matrix concatenating a $(n_k - n_{k-1}) \times 1$ vector of 1's horizontally with $\widehat{x}_k$ and $\widehat{Z}_k = I_3 \otimes \widehat{X}_k$, where $\otimes$ denotes the Kronecker product. The conditional posterior densities of ϕ_k and Σ_k are

$$\phi_k \sim N(b_k, B_k)$$

$$\Sigma_k \sim IW(\Omega_k, v_k)$$

where

$$B_k = \left(\widehat{Z}_k'(\Sigma_k^{-1} \otimes I_{n_k - n_{k-1}})\widehat{Z}_k + B_0^{-1}\right)^{-1}$$

$$b_k = B_k \left(\widehat{Z}_k'(\Sigma_k^{-1} \otimes I_{n_k - n_{k-1}})\widehat{y}_k + B_0^{-1} b_0 \right)$$

$$\Omega_k = \Omega_0 + \sum_{t=n_{k-1}+1}^{n_k} (y_t - \mu_k - \Phi_k y_{t-1})(y_t - \mu_k - \Phi_k y_{t-1})'$$

$$v_k = v_0 + (n_k - n_{k-1})$$

4. Sample b_0 from $p(b_0|Y_T, \Theta, \Theta_{0\_b_0}, S)$, where $\Theta_{0\_b_0}$ denotes the subset of Θ_0 excluding the parameter b_0.

$$b_0 \sim N(a, A)$$

where

$$A = \left(K B_0^{-1} + A_0^{-1}\right)^{-1}$$

$$a = A \left(B_0^{-1} \sum_{k=1}^{K} \phi_k + A_0^{-1} a_0 \right)$$

5. Sample B_0 from $p(B_0|Y_T, \Theta, \Theta_{0\_B_0}, S)$, where $\Theta_{0\_B_0}$ denotes the subset of Θ_0 excluding the parameter B_0.

$$B_0 \sim IW \left(D_0 + \sum_{k=1}^{K} (\phi_k - b_0)(\phi_k - b_0)', d_0 + K \right)$$

6. Sample Ω_0 from $p(\Omega_0|Y_T, \Theta, \Theta_{0\_\Omega_0}, S)$, where $\Theta_{0\_\Omega_0}$ denotes the subset of Θ_0 excluding the parameter Ω_0, (Metropolis step).

Since the conditional posterior density of Ω_0 is non-standard, a Metropolis-Hastings step is introduced to sample Ω_0. At iteration j, the proposal distribution for $\Omega_0^{(j)}$ is $IW\left((c_1 - m - 1)\Omega_0^{(j-1)}, c_1\right)$, which calibrates the mean of the proposal at $\Omega_0^{(j-1)}$. The free parameter c_1 controls the variation of the proposal: a bigger value of c_1 implies smaller variation and hence higher acceptance rate of the Metropolis step. m is the dimension of y_t. Note the prior density is $p(\Omega_0) = IW(\Omega_0|\Psi_0, f_0)$ and the likelihood function is $p(\Sigma_1, ..., \Sigma_K|\Omega_0, v_0) = \prod_{k=1}^{K} IW(\Sigma_k|\Omega_0, v_0)$. For any candidate draw Ω_0^* from the proposal distribution, the resulting acceptance probability is given by

$$\zeta_1 = \min \left\{ 1, \frac{IW(\Omega_0^*|\Psi_0, f_0) \prod_{k=1}^{K} IW(\Sigma_k|\Omega_0^*, v_0) IW\left(\Omega_0^{(j-1)}|(c_1 - m - 1)\Omega_0^*, c_1\right)}{IW(\Omega_0^{(j-1)}|\Psi_0, f_0) \prod_{k=1}^{K} IW(\Sigma_k|\Omega_0^{(j-1)}, v_0) IW\left(\Omega_0^*|(c_1 - m - 1)\Omega_0^{(j-1)}, c_1\right)} \right\}$$

With probability ζ_1, the candidate draw Ω_0^* is accepted and $\Omega_0^{(j)} = \Omega_0^*$. Otherwise, $\Omega_0^{(j)} = \Omega_0^{(j-1)}$.

7. Sample v_0 from $p(v_0|Y_T, \Theta, \Theta_{0\_v_0}, S)$, where $\Theta_{0\_v_0}$ denotes the subset of Θ_0 excluding the parameter v_0, (Metropolis step).

A Metropolis-Hastings step is used to sample v_0. At iteration j, the proposal distribution for $v_0^{(j)}$ is $Gamma\left(v_0^{(j-1)}/c_2, c_2\right)$, which calibrates the mean of the proposal at $v_0^{(j-1)}$. The free parameter c_2 controls the variation of the proposal: a bigger value of c_2 implies larger variation and hence lower acceptance rate of the Metropolis step. Note the prior density is $p(v_0) = Gamma(v_0|\rho_0, \lambda_0)$ and the likelihood function is $p(\Sigma_1, ..., \Sigma_K|\Omega_0, v_0) = \prod_{k=1}^K IW(\Sigma_k|\Omega_0, v_0)$. For any candidate draw v_0^* from the proposal distribution, the resulting acceptance probability is given by

$$\zeta_2 = \min\left\{1, \frac{Gamma(v_0^*|\rho_0, \lambda_0)\prod_{k=1}^K IW(\Sigma_k|\Omega_0, v_0^*)Gamma\left(v_0^{(j-1)}|v_0^*/c_2, c_2\right)}{Gamma(v_0^{(j-1)}|\rho_0, \lambda_0)\prod_{k=1}^K IW(\Sigma_k|\Omega_0, v_0^{(j-1)})Gamma\left(v_0^*|v_0^{(j-1)}/c_2, c_2\right)}\right\}$$

With probability ζ_2, the proposed draw v_0^* is accepted and $v_0^{(j)} = v_0^*$. Otherwise, $v_0^{(j)} = v_0^{(j-1)}$.

8. Sample α_0 from $p(\alpha_0|Y_T, \Theta, \Theta_{0\_\alpha_0}, S)$, where $\Theta_{0\_\alpha_0}$ denotes the subset of Θ_0 excluding the parameter α_0, (Metropolis step).

A Metropolis-Hastings step is used to sample α_0. At iteration j, the proposal distribution for $\alpha_0^{(j)}$ is $Gamma\left(\alpha_0^{(j-1)}/c_3, c_3\right)$, which calibrates the mean of the proposal at $\alpha_0^{(j-1)}$. The free parameter c_3 controls the variation of the proposal: a bigger value of c_3 implies larger variation and hence lower acceptance rate of the Metropolis step. Note the prior density is $p(\alpha_0) = Gamma(\alpha_0|q_0, \gamma_0)$ and the likelihood function is $p(\pi_1, ..., \pi_{K-1}|\alpha_0, \beta_0) = \prod_{k=1}^{K-1} Beta(\pi_k|\alpha_0, \beta_0)$. For any candidate draw α_0^* from the proposal distribution, the resulting acceptance probability is given by

$$\zeta_3 = \min\left\{1, \frac{Gamma(\alpha_0^*|q_0, \gamma_0)\prod_{k=1}^{K-1} Beta(\pi_k|\alpha_0^*, \beta_0)Gamma\left(\alpha_0^{(j-1)}|\alpha_0^*/c_3, c_3\right)}{Gamma(\alpha_0^{(j-1)}|q_0, \gamma_0)\prod_{k=1}^{K-1} Beta(\pi_k|\alpha_0^{(j-1)}, \beta_0)Gamma\left(\alpha_0^*|\alpha_0^{(j-1)}/c_3, c_3\right)}\right\}$$

With probability ζ_3, the proposed draw α_0^* is accepted and $\alpha_0^{(j)} = \alpha_0^*$. Otherwise, $\alpha_0^{(j)} = \alpha_0^{(j-1)}$.

9. Sample β_0 from $p(\beta_0|Y_T, \Theta, \Theta_{0\_\beta_0}, S)$, where $\Theta_{0\_\beta_0}$ denotes the subset of Θ_0 excluding the parameter β_0, (Metropolis step).

A Metropolis-Hastings step is used to sample β_0. At iteration j, the proposal distribution for $\beta_0^{(j)}$ is $Gamma\left(\beta_0^{(j-1)}/c_4, c_4\right)$, which calibrates the mean of the proposal at $\beta_0^{(j-1)}$. The free parameter c_4 controls the variation of the proposal: a bigger value of c_4 implies larger variation and hence lower acceptance rate of the Metropolis step. Note the prior density is $p(\beta_0) = Gamma(\beta_0|r_0, \delta_0)$ and the likelihood function is $p(\pi_1, ..., \pi_{K-1}|\alpha_0, \beta_0) = \prod_{k=1}^{K-1} Beta(\pi_k|\alpha_0, \beta_0)$. For any candidate draw β_0^* from the proposal distribution, the resulting acceptance probability is given by

$$\zeta_4 = \min\left\{1, \frac{Gamma(\beta_0^*|r_0, \delta_0)\prod_{k=1}^{K-1} Beta(\pi_k|\alpha_0, \beta_0^*)Gamma\left(\beta_0^{(j-1)}|\beta_0^*/c_4, c_4\right)}{Gamma(\beta_0^{(j-1)}|r_0, \delta_0)\prod_{k=1}^{K-1} Beta(\pi_k|\alpha_0, \beta_0^{(j-1)})Gamma\left(\beta_0^*|\beta_0^{(j-1)}/c_4, c_4\right)}\right\}$$

With probability ζ_4, the proposed draw β_0^* is accepted and $\beta_0^{(j)} = \beta_0^*$. Otherwise, $\beta_0^{(j)} = \beta_0^{(j-1)}$.

This completes the algorithm for sampling from the structural break VAR model.

A.2 Marginal Likelihood of the Structural Break VAR Model

In this paper, we adopt the modified harmonic mean (MHM) method of Gelfand and Dey (1994) to compute the marginal likelihood. The basic idea of the MHM method is to utilize the simple identity which is obtained by rearranging Bayes rule.

$$\frac{1}{p(Y_t|s_t = K)} = \frac{p(\Theta, \Theta_0|Y_t, s_t = K)}{p(\Theta, \Theta_0|s_t = K)p(Y_t|\Theta, \Theta_0, s_t = K)}$$

Note the priors of Θ and Θ_0 are independent of the number of in-sample regimes. So $p(\Theta, \Theta_0|s_t = K) = p(\Theta_0)p(\Theta|\Theta_0)$. Also the likelihood of Y_t does not depend on Θ_0 given

the value of Θ. So $p(Y_t|\Theta, \Theta_0, s_t = K) = p(Y_t|\Theta, s_t = K)$.

For any density function $h(\Theta, \Theta_0)$ whose support is contained in that of $p(\Theta, \Theta_0|Y_t, s_t = K)$, it is easy to show that

$$\frac{1}{p(Y_t|s_t = K)} = \int \frac{h(\Theta, \Theta_0)}{p(\Theta_0)p(\Theta|\Theta_0)p(Y_t|\Theta, s_t = K)} p(\Theta, \Theta_0|Y_t, s_t = K) d\Theta d\Theta_0$$

Given a sample of posterior draws $\{\Theta^{(j)}, \Theta_0^{(j)}\}_{j=1}^n$ from $p(\Theta, \Theta_0|Y_t, s_t = K)$, the log marginal likelihood can be computed as

$$\log\left(p(Y_t|s_t = K)\right) \approx -\log\left(\frac{1}{n}\sum_{j=1}^{n} \frac{h(\Theta^{(j)}, \Theta_0^{(j)})}{p(\Theta_0^{(j)})p(\Theta^{(j)}|\Theta_0^{(j)})p(Y_t|\Theta^{(j)}, s_t = K)}\right)$$

The likelihood function $p(Y_t|\Theta, s_t = K)$ can be obtained as a by-product from the Hamilton filter in the estimation process.

One condition for this numerical integration is that the function $\frac{h(\Theta,\Theta_0)}{p(\Theta_0)p(\Theta|\Theta_0)p(Y_t|\Theta,s_t=K)}$ needs to be bounded above for the rate of convergence to be practical (Geweke(1999)). Geweke (1999) proposes a convenient implementation of $h(\cdot)$ which satisfies the above condition. The function $h(\cdot)$ is chosen to be a truncated Gaussian density with the mean and covariance matrix constructed from posterior draws of $p(\Theta, \Theta_0|Y_t, s_t = K)$. Specifically, let $\widetilde{\Theta} = \{\Theta, \Theta_0\}$, $\overline{\Theta} = \frac{1}{n}\sum_{j=1}^{n} \widetilde{\Theta}^{(j)}$ and $\overline{\Omega} = \frac{1}{n}\sum_{j=1}^{n} \left(\widetilde{\Theta}^{(j)} - \overline{\Theta}\right)\left(\widetilde{\Theta}^{(j)} - \overline{\Theta}\right)'$. Define $\widehat{\Theta} \equiv \left\{\widetilde{\Theta} : \left(\widetilde{\Theta} - \overline{\Theta}\right)' \overline{\Omega}^{-1} \left(\widetilde{\Theta} - \overline{\Theta}\right) \leq \chi_p^2(m)\right\}$, where m is the dimension of $\widetilde{\Theta}$ and $\chi_p^2(m)$ is the $p \times 100$ percent critical value of a χ^2 distribution with m degrees of freedom. The implementation of $h(\cdot)$ is given by

$$h\left(\widetilde{\Theta}\right) = \frac{1}{p(2\pi)^{0.5m}\left|\overline{\Omega}\right|^{0.5}} \exp\left(-0.5\left(\widetilde{\Theta} - \overline{\Theta}\right)' \overline{\Omega}^{-1} \left(\widetilde{\Theta} - \overline{\Theta}\right)\right) I_{\widehat{\Theta}}$$

where

$$I_{\widehat{\Theta}} = \begin{cases} 1, & \text{if } \widetilde{\Theta} \in \widehat{\Theta}; \\ 0, & \text{otherwise.} \end{cases}$$

A smaller value of the truncation probability p will likely result in better behavior of the numerical integration since more tail draws are discarded. But greater simulation error may occur as fewer draws are retained in the set $\widehat{\Theta}$. In practice, it is usually chosen to be in the range $(0.9, 1)$.

Bibliography

Ang, A., and G. Bekaert (2002): "Regime Switches in Interest Rates," *Journal of Business and Economic Statistics*, 20, 163–182.

Ang, A., M. Piazzesi, and M. Wei (2006): "What Does the Yield Curve Tell Us About GDP Growth?," *Journal of Econometrics*, 131, 359–403.

Ball, C., and A. Roma (1994): "Stochastic Volatility Option Pricing," *Journal of Financial and Quantitative Analysis*, 29, 589–607.

Bansal, R., and H. Zhou (2002): "Term Structure of Interest Rates with Regime Shifts," *Journal of Finance*, 57, 1997–2043.

Barry, D., and J. Hartigan (1993): "A Bayesian analysis for change point problems," *Journal of the American Statistical Association*, 88, 309–319.

Bauwens, L., A. Preminger, and J. Rombouts (2006): "Regime Switching GARCH Models," Working Paper, University of Montreal.

Black, F. (1976): "Studies of Stock Market Volatility Changes," *Proceedings of the American Statistical Association, Business and Economic Statistic Section*, pp. 177–181.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 309–328.

Carpenter, J., P. Clifford, and P. Fearnhead (1999): "An improved particle filter for nonlinear problems," *IEE Proceedings-Radar, Sonar and Navigation*, 146, 2–7.

Carvalho, C., and H. Lopes (2007): "Simulation-based sequential analysis of Markov switching stochastic volatility models," *Computational Statistics and Data Analysis*, 51, 4526–4542.

Casarin, R., and J.-M. Marin (2007): "Online data processing: comparison of Bayesian regularized particle filters," University of Brescia, Department of Economics, working paper n. 0704.

Casarin, R., and C. Trecroci (2006): "Business Cycle and Stock Market Volatility: A Particle Filter Approach," Cahier du CEREMADE N. 0610, University Paris Dauphine.

Chen, N. (1991): "Financial Investment Opportunities and the Macroeconomy," *Journal of Finance*, 46, 529–554.

Chib, S. (1995): "Marginal Likelihood from the Gibbs Sampler," *Journal of the American Statistical Association*, 90, 1313–1321.

Chib, S. (1996): "Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models," *Journal of Econometrics*, 75, 79–97.

——— (1998): "Estimation and Comparison of Multiple Change Point Models," *Journal of Econometrics*, 86, 221–241.

——— (2001): "Markov Chain Monte Carlo Methods: Computation and Inference," in *Handbook of Econometrics*, ed. by Heckman, and Leamer. Elsevier Science.

Chib, S., and I. Jeliazkov (2001): "Marginal Likelihood from the Metropolis-Hasting Output," *Journal of the American Statistical Association*, 96, 270–281.

Chopin, N. (2007): "Dynamic detection of change points in long time series," *Annals of the Institute of Statistical Mathematics*, 59, 349–366.

Chopin, N., and F. Pelgrin (2004): "Bayesian inference and state number determination for hidden Markov models: an application of the information content of the yield curve about inflation," *Journal of Econometrics*, 123(2), 327–344.

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

Chu, J. (1995): "Detecting parameter shift in GARCH models," *Econometric Reviews*, 14, 241–266.

Danielsson, J., and J. Richard (1993): "Accelerated Gaussian Importance Sampler with Applications to Dynamic Latent Variable Models," *Journal of Applied Econometrics*, 8, 153–173.

Davis, P., and G. Fagan (1997): "Are Financial Spreads Useful Indicators of Future Inflation and Output Growth in EU Countries?," *Journal of Applied Econometrics*, 12, 701–714.

Diebold, F. (1986): "Modeling the persistence of conditional variances: a comment," *Econometric Reviews*, 5, 51–56.

Doucet, A., N. de Freitas, and N. Gordon (eds.) (2001): *Sequential Monte Carlo Methods in Practice*. Springer-Verlag.

Durbin, J., and S. Koopmans (1997): "Monte Carlo Maximum Likelihood Estimation for Non-Gaussian State Space Models," *Biometrica*, 84, 669–684.

Engle, R. F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the UK inflation," *Econometrica*, 50, 987–1008.

Estrella, A., and G. Hardouvelis (1991): "The Term Structure as a Predictor of Real Economic Activity," *Journal of Finance*, 46, 555–576.

Estrella, A., and F. Mishkin (1997): "The Predictive Power of the Term Structure of Interest Rates in Europe and the United States: Implications for the European Central Bank," *European Economic Review*, 41, 1375–1401.

——— (1998): "Predicting U.S. Recessions: Financial Variables as Leading Indicators," *Review of Economics and Statistics*, 80, 45–61.

Estrella, A., A. Rodrigues, and S. Schich (2003): "How Stable is the Predictive Power of the Yield Curve? Evidence from Germany and the United States," *Review of Economics and Statistics*, 85, 629–644.

Fama, E., and R. Bliss (1987): "The Information in Long-Maturity Forward Rates," *American Economic Review*, 77, 680–692.

French, K., and R. Roll (1986): "Stock Return Variances: The Arrival of Information and The Reaction of Traders," *Journal of Financial Economics*, 17, 5–26.

French, K., G. Schwert, and R. Stambaugh (1987): "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3–29.

Fridman, M., and L. Harris (1998): "A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models," *Journal of Business and Economics Statistics*, 16, 284–291.

Fruhwirth-Schnatter, S. (1995): "Bayesian Model Discrimination and Bayes Factor for Linear Gaussian State Space Models," *Journal of the Royal Statistical Society, Series B*, 57, 237–246.

——— (2004): "Estimating Marginal Likelihoods for Mixture and Markov Switching Models Using Bridge Sampling Techniques," *The Econometrics Journal*, 7, 143–167.

Gelfand, A., and D. Dey (1994): "Bayesian Model Choice: Asymptotics and Exact Calculations," *Journal of The Royal Statistical Society*, B, 56, 501–514.

Gerlach, R., C. Carter, and R. Kohn (2000): "Efficient Bayesian inference for dynamic mixture models," *Journal of the American Statistical Association*, 88, 819–828.

Geweke, J. (1989): "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317–1339.

——— (1994): "Bayesian Comparison of Econometric Models," Working Paper, Federal Reserve Bank of Minneapolis.

Geweke, J. (2005): *Contemporary Bayesian Econometrics and Statistics*. John Wiley and Sons Ltd.

Geweke, J., and C. Whiteman (2005): "Bayesian Forecasting," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. Granger, and A. Timmermann, vol. forthcoming.

Ghysels, E., A. Harvey, and E. Renault (1996): "Stochastic Volatility," in *Statistical Methods in Finance, Handbook of Statistics*, ed. by G. Maddala, and C. Rao, vol. 14. North-Holland, Amsterdam.

Giacomini, R., and B. Rossi (2006): "How Stable is the Forecasting Performance of the Yield Curve for Output Growth?," *Oxford Bulletin of Economics and Statistics*, 68, 783–795.

Gilks, W. R., and C. Berzuini (2001): "Following a Moving Target-Monte Carlo Inference for Dynamic Bayesian Models," *Journal of the Royal Statistical Society. Series B*, 63(1), 127–146.

Giordani, P., and R. Kohn (2006): "Efficient Bayesian Inference for Multiple Change-Point and Mixture Innovation Models," Sveriges Riksbank Working Paper 196.

Gordon, N., D. Salmond, and A. Smith (1993): "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings*, F-140, 107–113.

Gourieroux, C., and A. Monfort (1996): *Simulation Based Econometric Methods*. Oxford University Press.

Gray, S. F. (1996): "Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process," *Journal of Financial Economics*, 42, 27–62.

Green, P. J. (1995): "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82, 771–732.

Haas, M., S. Mittnik, and M. Paolella (2004): "A New Approach to Markov Switching GARCH Models," *Journal of Financial Econometrics*, 2, 493–530.

Hamilton, J., and D. Kim (2002): "A Re-Examination of the Predictability of Economic Activity using the Yield Spread," *Journal of Money, Credit and Banking*, 34, 340–360.

Hamilton, J. D. (1988): "Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates," *Journal of Economic Dynamics and Control*, 12, 385–423.

Hamilton, J. D., and R. Susmel (1994): "Autoregressive Conditional Heteroskedasticity and Changes in Regime," *Journal of Econometrics*, 64, 307–333.

Hansson, B., and P. Hordahl (2005): "Forecasting Variance Using Stochastic Volatility and GARCH," *The European Journal of Finance*, 11, 33–57.

Harvey, A., and N. Shephard (1996): "Estimation of An Asymmetric Stochastic Volatility Model for Asset Returns," *Journal of Business and Economic Statistics*, 14, 429–434.

Harvey, C. R. (1989): "The Real Term Structure and Consumption Growth," *Journal of Financial Economics*, 22, 305–333.

He, Z., and J. Maheu (2008): "Real Time Detection of Structural Breaks in GARCH Models," Working Paper 336, Department of Economics, University of Toronto.

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

Hillebrand, E. (2005): "Neglecting parameter changes in GARCH models," *Journal of Econometrics*, 127, 121–138.

Hull, J., and A. White (1987): "The Pricing of Options on Assets with Stochastic Volatility," *Journal of Finance*, 42, 381–400.

Jacquier, E., N. Polson, and P. Rossi (2004): "Bayesian Analysis of Stochastic Volatility Models with Fat-tails and Correlated Errors," *Journal of Econometrics*, 122, 185–212.

Johannes, M., and N. Polson (2006): "Exact Bayesian particle filtering and parameter learning," Working Paper, University of Chicago.

Kadiyala, R., and S. Karlsson (1997): "Numerical Methods for Estimation and Inference in Bayesian VAR Models," *Journal of Applied Econometrics*, 12, 99–132.

Kass, R. E., and A. E. Raftery (1995): "Bayes Factors," *Journal of the American Statistical Association*, 90(420), 773–795.

Kim, C., J. Morley, and C. Nelson (2005): "The structural breaks in the equity premium," *Journal of Business and Economic Statistics*, 23, 181–191.

Kim, S., N. Shephard, and S. Chib (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361–393.

Kitagawa, G. (1996): "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," *Journal of Computational and Graphical Statistics*, 5, 1–25.

Klaassen, F. (2002): "Improving GARCH volatility forecasts with regime switching GARCH," *Empirical Economics*, 27, 363–394.

Koop, G. (2003): *Bayesian Econometrics*. Wiley, Chichester, England.

Koop, G., and S. Potter (2007): "Estimation and Forecasting in Models with Multiple Breaks," *Review of Economic Studies*, 74, 763–789.

Koopman, S., N. Shephard, and D. Creal (2009): "Testing The Assumptions Behind Importance Sampling," *Journal of Econometrics*, 149, 2–11.

Lamoureux, G. C., and W. D. Lastrapes (1990): "Persistence in Variance, Structural Change, and the GARCH Model," *Journal of Business & Economic Statistics*, 8, 225–234.

Laurent, R. (1988): "An Interest Rate-Based Indiator of Monetary Policy," *Federal Reserve Bank of Chicago Economic Perspectives*, 12, 3–14.

——— (1989): "Testing the Spread," *Federal Reserve Bank of Chicago Economic Perspectives*, 13, 22–34.

Litterman, R. (1980): "A Bayesian Procedure for Forecasting with Vector Autoregression," Mimeo, Massachusetts Institute of Technology.

——— (1986): "Forecasting with Bayesian Vector Autoregression: Five Years of Experience," *Journal of Business and Economic Statistics*, 4, 25–38.

Liu, C., and J. Maheu (2008): "Are There Structural Breaks in Realized Volatility ?," *Journal of Financial Econometrics*, 6, 326–360.

Liu, J., and R. Chen (1998): "Sequential Monte Carlo methods for dynamical systems," *Journal of the American Statistical Association*, 93, 1032–1044.

Liu, J., and M. West (2001): "Combined parameter and state estimation in simulation-based filtering," in *Sequential Monte Carlo Methods in Practice*, ed. by A. Doucet, N. de Freitas, and N. Gordon. Springer-Verlag.

Lutkepohl, H. (2006): *New Introduction to Multiple Time Series Analysis*. Springer.

Maheu, J., and T. McCurdy (2009): "How Useful are Historical Data for Forecasting the Long-Run Equity Return Distribution ?," *Journal of Business and Economic Statistics*, 27, 95–112.

Maheu, J. M., and S. Gordon (2008): "Learning, Forecasting and Structural Breaks," *Journal of Applied Econometrics*, 23, 553–583.

McCulloch, R., and R. Tsay (1993): "Bayesian inference and prediction for mean and variance shifts in autoregressive time series," *Journal of the American Statistical Association*, 88, 968–978.

Meng, X., and W. Wong (1996): "Simulating Ratios of Normalizing Constants via a Simple Identity," *Statistical Sinica*, 6, 831–860.

Miazhynskaia, T., and G. Dorffner (2006): "A Comparison of Bayesian Model Selection Based on MCMC with an Application to GARCH-Type Models," *Statistical Papers*, 47, 525–549.

Mikosch, T., and C. Starica (2004): "Nonstationarities in financial time series, the long-range dependence and the IGARCH effects," *Review of Economics and Statistics*, 86, 378–390.

Nelson, D. B. (1991): "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

Newey, W., and K. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Newton, M. A., and A. Raftery (1994): "Approximate Bayesian inference by the weighted likelihood bootstrap (with Discussion).," *Journal of The Royal Statistical Society*, B, 56, 3–48.

Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007): "Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference," *Journal of Econometrics*, 140, 425–449.

Owen, A., and Y. Zhou (2000): "Safe and Effective Importance Sampling," *Journal of the American Statistical Association*, 95, 135–143.

Pastor, L., and R. F. Stambaugh (2001): "The Equity Premium and Structural Breaks," *Journal of Finance*, 4, 1207–1231.

Pesaran, H., D. Pettenuzzo, and A. Timmermann (2006): "Forecasting Time Series Subject to Multiple Structural Breaks," *Review of Economic Studies*, 73, 1057–1084.

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

Plosser, C., and G. Rouwenhorst (1994): "International Term Structure and Real Economic Growth," *Journal of Monetary Economics*, 33, 133–156.

Polson, N., J. Stroud, and P. Muller (2008): "Practical Filtering with Sequential Parameter Learning," *Journal of the Royal Statistical Society: Series B*, 70, 413–428.

Raggi, D., and S. Bordignon (2008): "Sequential Monte Carlo Methods for Stochastic Volatility Models with Jumps," working paper, Department of Economics, University of Bologna.

Rapach, D., and J. Strauss (2006): "Structural breaks and GARCH models of exchange rate volatility," *Journal of Applied Econometrics*, forthcoming.

Richard, J., and W. Zhang (2007): "Efficient High-Dimensional Importance Sampling," *Journal of Econometrics*, 141, 1385–1411.

Ristic, B., S. Arulampalam, and N. Gordon (eds.) (2004): *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, Boston.

Robert, C., and G. Casella (2004): *Monte Carlo Statistical Methods*. Springer, New York.

Schwert, G. (1989): "Why Does Stock Market Volatility Change Over Time?," *Journal of Finance*, 44, 1115–1154.

Scott, S. L. (2002): "Bayesian Methods for Hidden Markov Models: Recursive Computing in the 21st Century," *Journal of the American Statistical Association*, 97(457), 337–351.

Shephard, N. (ed.) (2005): *Stochastic Volatility: Selected Readings*. Oxford University Press.

Starica, C., and C. Granger (2005): "Nonstationarities in Stock Returns," *Review of Economics and Statistics*, 87(3), 503–522.

Stock, J., and M. Watson (1989): "New Indexes of Coincident and Leading Economic Indicators," in *NBER Macroeconomics Annual*, ed. by O. Blanchard, and S. Fisher, pp. 352–394.

Stock, J. H., and M. W. Watson (1999): "Forecasting Inflation," *Journal of Monetary Economics*, 44, 293–335.

Stock, J. H., and M. W. Watson (2003): "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, 41(3), 788–829.

Storvik, G. (2002): "Particle Filters in State Space Models With the Presence of Unknown Static Parameters," *IEEE Transactions Signal Processing*, 50, 281–289.

Taylor, S. J. (1986): *Modeling Financial Time Series*. John Wiley.

Tierney, L., and J. Kadane (1986): "Accurate Approximations for Posterior Moments and Marginal Densities," *Journal of the American Statistical Association*, 81, 82–86.

Yu, J. (2005): "On Leverage In a Stochastic Volatility Model," *Journal of Econometrics*, 127, 165–178.