Essays on Volatility Models Using EMM Estimation

Ying Gu

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2006

Program Authorized to Offer Degree:
Department of Economics

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

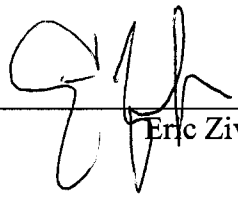
University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Ying Gu

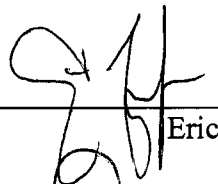
and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of the Supervisory Committee:



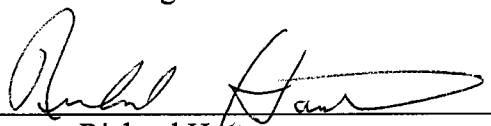
Eric Zivot

Reading Committee:



Eric Zivot

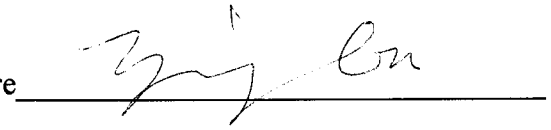
Doug Martin



Richard Hartman

Date: 7/18/2006

In presenting this dissertation in partial fulfillment of the requirements for the doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of the dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106-1346, 1-800-521-0600, or to the author.

Signature 

Date 7/20/2006

University of Washington

Abstract

Essays on Volatility Models Using EMM Estimation

Ying Gu

Chair of the Supervisory Committee:
Associate Professor Eric Zivot
Department of Economics

Academic researchers and investment institutions have devoted a significant amount of their efforts over the past two decades to developing and testing sophisticated models of the volatility dynamics of different types of asset-pricing data. In a set of three essays, I analyze the various aspects of model specification issues by employing the efficient method of moments (EMM) technique. After an introduction to EMM methodology and procedure, my first essay offers a comprehensive comparison of *univariate* volatility models for US short rates. We find that a continuous-time two-factor SV model, a continuous-time three-factor SV model, and a discrete-time RS-in-volatility model with level effect can well explain the salient features of the short rate. We also show that either an SV model with a level effect or a RS model with a level effect, but not both, is needed for explaining the data. Our EMM estimates of the level effect are much lower than unity, but around 1/2 after incorporating the SV effect or the RS effect. The second essay applies appropriate filtering and smoothing algorithms on the simulated data from the preferred volatility models and on the series of the US short rate. The third essay aims to carry out a Monte Carlo experiment for estimating a Markov RS model using EMM. This dissertation spans the fields of Econometrics, Finance and Macroeconomics; it provides crucial insights for both academics and practitioners due to its implications for macroeconomic policy and for microeconomic decisions in the theory and practice of asset pricing, asset allocation, and risk management.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

LIST OF TABLES

Table Number Page

ACKNOWLEDGEMENTS

I can not find enough words to express my gratitude to my adviser Eric Zivot for his continuous advice, help, and support. This dissertation could not have been written in its current form without him. I also thank my other committee members, Professor Doug Martin and Professor Richard Hartman, for constant encouragement and an open-door policy for my questions. I thank the Grover and Creta Ensley Fellowship in Public Policy for financial support during my final year. My thanks also go to Jiahui Wang, George Tauchen, Hao Zhou, John Nolan, Angelo Melino, Bruce Wang, as well as seminar participants at University of Washington and conference participants at the 22nd Canadian Econometrics Study Group Conference, for helpful comments on Chapter 1.

DEDICATION

To my parents and Simon.

CHAPTER I: A Comparison of Univariate Stochastic Volatility Models using EMM Estimation for U.S. Short Rates

1.1. Introduction

The risk-free short-term interest rate is a key state variable in asset pricing models, term structure models and macroeconomic models. It is used to express the expected equilibrium returns on risky assets in terms of excess. It directly affects the short end of the term structure and thus has implications for the pricing of the full range of fixed income securities and derivatives. Further, the short rate is an important input for business cycle analysis through its impact on the cost of credit, its sensitivity to the stance of monetary policy, and to inflationary expectations.[1]

Originating from the Brownian motion representation of Merton (1973), an enormous amount of work has been directed towards modeling and estimating the dynamics of the short rate. The mean-reverting model in Vasicek (1977) allows the dynamics of interest rates to be stationary. The square-root model of Cox, Ingersoll, and Ross (1985) (CIR) guarantees positive interest rates and incorporates the "level effect," which allows volatility to increase with the level of the interest rate. Chan et al. (1992) (hereafter CKLS) compared a variety of single factor linear diffusion models for the short rate. They found that models that freely estimated the level effect outperformed other models, and that the level effect parameter estimate was significantly greater than unity. Due to the poor empirical performance of linear diffusion models, several authors have focused on the estimation of the functional form of the drift and volatility of the diffusion model. Conley et al. (1995) exploited the

[1] See Andersen (2005).

moment generating techniques of Hansen and Scheinkman (1995) to obtain nonparametric estimates of the drift; Ait-Sahalia (1996a) estimated the volatility function nonparametrically; and Stanton (1997) provided nonparametric discrete-time approximations to the drift and volatility functions.

The poor performance of one factor models led to the incorporation of an additional stochastic volatility (SV) factor in order to accommodate the strong conditional heteroskedasticity in short rates. Longstaff and Schwartz (1992) derived a two-factor general equilibrium model for the short rate, with its level and its conditional volatility as factors. They showed that a two-factor model improves upon a single factor model, and carries additional information about the term structure and leads to better pricing and hedging performance compared with a single factor model. Similarly, Brenner et al. (1996) and Koedijk et al. (1997) modeled the conditional volatility process of the short rate as a GARCH process and found that a model with both level and GARCH effects outperforms models that exclude one of them. Later, Anderson and Lund (1997) (hereafter AL) and Ball and Torous (1999) found that a two-factor model with level and SV factors outperforms the two-factor model with GARCH volatility.

In the specification of interest rate models, evidence has been documented for regime switching (RS) behavior in short rates. Garcia and Perron (1996) provided a three-regime model using the methodology of Hamilton (1989), which allows the drift and volatility of the ex-post real interest rate to switch over regimes. Their results suggested that both the drift and volatility are essentially different for the periods 1961-1973, 1973-1980 and 1980-1986. Cai (1994) presented a RS-ARCH model for the

excess returns of the three-month T-bill over the thirty-day T-bill and reported two periods of high interest rate volatility: one is in 1974 (the energy crisis) and the other is between 1979 and 1982 (the "monetary experiment" of the Federal Reserve). Gray (1996) developed a generalized RS model based on a CIR process with regime dependence in both mean reversion and conditional volatility driven by a GARCH process. He found evidence of a high (low) volatility regime with high (low) mean reversion for one-month U.S. T-Bill yields. An additional high-volatility regime is found in 1987, corresponding to the stock market crash. Gray argued that the RS *and* GARCH effect as well as the diffusion terms are necessary for accommodating the dynamics of the short rates. Smith (2002) presented a model for the short rate based on the CKLS process but only allowed the unconditional volatility to switch between regimes. He empirically compared the RS models and SV models using a quasi-maximum likelihood estimation technique, and argued that either a RS or an SV effect, but not both, is needed to adequately describe the data. Ang and Bekaert (2002a) found that regime-switching models of interest rates replicate non-linear patterns in the drift and volatility functions of short rates found in non-parametric approaches.

As pointed out by many authors, the RS model is more than a mere device used to fit the data; it has important implications for business cycle analysis and yield curve dynamics due to the natural association between the notion of regimes that underlie the econometric model and the large economy-wide shocks that have strong and persistent influences on the behavior of interest rates. For example, Naik and Lee (1998) showed that the RS model generates an empirically more reasonable term structure of volatilities, fat tails, and persistence in volatility compared to those of the SV models.

Ang and Bekaert (2001) argued that the two-regime classification of U.S. nominal short term rates corresponds reasonably well with business cycles. Lahiri et al. (2000) studied the comparative performance of a number of interest rate spreads as predictors of the German inflation and business cycle in the post-Bretton Woods era using a two-regime RS model. Bansal and Zhou (2002) developed a term structure model, compared the two-factor RS model with the benchmark CIR model and affine models with up to three factors, and argued that only the RS model can account for the well documented violations of the expectations hypothesis, the observed conditional volatility, and the conditional correlation across yields with regimes intimately related to business cycles. Dai, Singleton, and Yang (2004) developed a term structure model with priced factor and RS risks, provided closed-form solutions for zero-coupon bond prices, and argued that the shapes of the term structures of bond yield volatilities are very different across regimes.

In response to the non-Gaussian behavior of interest rates and asset returns, models have been developed that relax the assumption of conditionally normally distributed innovations to take into account of both volatility clustering and leptokurtosis in describing the financial series.. The GARCH model with Student-t distributed innovations was considered by Bollerslev (1987), and the GARCH model with the extended skewed Student-t distribution was utilized by Lambert and Laurent (2000). Other distributions have been examined, including the normal inverse Gaussian process by Barndorff-Nielsen (1997) and Andersson (2001), the variance-gamma process by Madan and Seneta (1990), the generalized hyperbolic process of Eberlein, Keller and Prause (1998), and the CGMY process by Carr, Geman, Madan and Yor

(2000). In general, estimates of conditional volatility using non-Gaussian distribution showed better results relative to estimates obtained assuming normality. For a review of these results see Peters (2001), and Verhoven and McAleer (2003).

While the SV model and its extensions have theoretical appeal, efficient estimation is not straightforward. Standard statistical methods, both classical and Bayesian, are usually not applicable either because it is not practicable to obtain the likelihood for the entire state vector or because the integration required to eliminate unobservable factors from the likelihood is infeasible. A variety of estimation procedures has been proposed to overcome these difficulties, including the generalized method of moments (GMM) used by Melino and Turnbull (1990), the quasi maximum likelihood (QML) approach followed by Harvey et al. (1994), the simulated maximum likelihood approaches used by Danielsson (1994) and Sandmann and Koopman (1998), the Markov-chain Monte Carlo (MCMC) procedures used by Jaquier et al. (1994) and Kim et al. (1998), and the efficient methods of moments (EMM) approach developed by Gallant and Tauchen (1996) and Gallant and Long (1997). [2]

Although there is a large literature on SV models for interest rates, there still remains substantial disagreement on the empirical performance of different model specifications. The main reason for these disagreements is the use of estimation techniques that make it difficult to compare competing models in a unified way. In this paper we follow the methodology of Gallant, Hsieh, and Tauchen (1997) and use the EMM to estimate and compare a comprehensive collection of univariate SV models for

[2] See Andersen et al. (1990) for performance comparisons, and Broto and Ruiz (2002) for a survey on asymptotic properties, finite sample experiments, limitations and advantages of various estimators. Shephard (2005) provides a general overview of the literature.

the short-term interest rate including one-factor diffusion models, two-factor and three-factor stochastic volatility (SV) models, non-Gaussian diffusion models with stable distributed errors, and a variety of Markov regime switching (RS) models. The use of EMM allows for a straightforward comparison of models, even if the models are non-nested.

Our results favor the one-factor non-Gaussian diffusion model over the one-factor Gaussian diffusion model, and the multi-factor SV models and the RS models over the one-factor non-Gaussian diffusion model. We show that a two-factor SV model, a three-factor SV model, and a RS-in-volatility model that allows for a level effect adequately describe the salient features of the short rate process. Our results show that the EMM estimates of the level effect are much lower than unity in the accepted SV models and RS model. Specifically, in our two-factor SV and three-factor SV models, the level effects are estimated similarly to that found in other studies of two-factor models (e.g. AL). In addition, the level effect estimate obtained from our RS-in-volatility model is also found to be around 1/2. Finally, we provide the first EMM estimations for a series of forms of the RS models and offer a performance comparison between different RS models and between the RS models and SV models for fitting the U.S. short rates. Our EMM estimation results clearly indicate that either an SV effect or a RS effect, but not both, are needed for describing the data accurately.

The remainder of the paper is organized as follows. Section 2 provides a description of the EMM methodology, procedure, and diagnostics. Section 3 presents the models for the short rate to be estimated and compared. Section 4 describes the data,

and Section 5 reports the EMM estimation results and processes the diagnostic tests. Section 6 summarizes and concludes.

1.2. Methodology

To facilitate a consistent evaluation and estimation across non-nested models, we rely on the EMM estimation technique developed in Gallant and Tauchen (1996) and extended in Gallant and Long (1997). The basic procedure of EMM estimation, summarized in Figure 1.1, consists of two steps.[3] First, in the projection step, the empirical conditional density of the observed time series is estimated by a semi-nonparametric (SNP) series expansion. This SNP expansion has a VAR-GARCH Gaussian density as its leading term, and departures from the Gaussian leading term are captured by a Hermite polynomial expansion. Second, in the estimation step, a GMM-type criterion function is constructed using the score functions from the log-likelihood of the SNP density as moments. The scores are evaluated using simulated data from a given structural model, and the criterion function is minimized with respect to the parameters underlying the structural model. A brief description of these steps, following Gallant and Tauchen (2001), is given below.

1.2.1 Projection Step

Gallant and Tauchen (2001) recommended the SNP model as the score generator for use with the EMM estimation. The advantage of the SNP model is that it can

[3]See Bansal and Zhou (2002).

approximate virtually any smooth distribution, even a mixture distribution (as is the case with a model of regime shifts).

To describe the SNP model, let y_t denote the observed data, and let $x_{t-1} = \{y_{t-1}, ..., y_1\}$ denote the lagged observations representing the complete and relevant information set. A SNP model starts with a Gaussian vector autoregression (VAR) with L_u lags, and a GARCH (L_g , L_r) or ARCH (L_r) conditional variance specification. The innovation density is a Hermite density of degree K_z, having the form of a polynomial times the standard normal density.

The SNP conditional density, $f(y_t \mid x_{t-1}, \theta)$, with parameter vector θ, has the form:

$$f(y_t \mid x_{t-1}, \theta) \propto \left[P(z_t)\right]^2 N(y_t \mid \mu_x, \Sigma_x) \tag{1.1}$$

where $z = R_x^{-1}(y - \mu_x)$ with $\Sigma_x = R_x R_x'$. $N(.)$ is a normal density of y with conditional mean μ_x and conditional variance Σ_x , where μ_x is estimated using a VAR specification, and Σ_x is estimated using an ARCH/GARCH specification, which parameterizes R_x.

To accommodate any remaining non-Gaussianity and time series structure in the innovation process, P is a Hermite polynomial with degree K_z in z ; to allow for additional conditional heterogeneity over that allowed by GARCH, the coefficients of the polynomial in the Hermite density are themselves polynomials of degree K_x in L_p

lags of the data.[4] For example, if only x_{t-1} is allowed to impact the conditional distribution, the Hermite polynomial P is given by

$$P(z_t, x_{t-1}) = \sum_{i=0}^{K_z} a_i z_t^i$$

where

$$a_i(x_{t-1}) = \sum_{j=0}^{K_x} a_{ij} x_{t-1}^j \tag{1.2}$$

The order of the polynomial expansion, K_z, controls the extent to which the tails deviate from normality. If $K_z = 0$, the SNP reduces to the normal density. The order of the coefficients of the polynomial, K_x, determines the degree of the heterogeneity of the innovations $\{z_t\}$. When $K_x = 0$, z_t are homogeneous, that is, the conditional density is independent of the lagged observations, x_{t-1}. If $K_x > 0$, we effectively multiply the innovations by functions of x_{t-1}.

Because the number of terms in a polynomial expansion becomes exponentially large as the dimension increases, two additional tuning parameters are introduced: $I_z > 0$ implies that all interactions larger than $K_z - I_z$ are suppressed; similarly for $I_x > 0$. The tuning parameters that describe a SNP model are summarized by the vector $(L_u, L_g, L_r, L_p, K_z, I_z, K_x, I_x)$. Table 1.1 gives a taxonomy of common SNP models.

For a given set of set of tuning parameters, the parameters θ of the SNP model are estimated by quasi-maximum likelihood (QML). The quasi-maximum likelihood estimator, $\tilde{\theta}_n$ satisfies the first-order conditions of the optimization problem,

---

[4] See Gallant and Tauchen (1996).

$$m_n(\tilde{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\ln f(y_t \mid x_{t-1}, \tilde{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n} s_f(y_t \mid x_{t-1}, \tilde{\theta}_n) = 0 \qquad (1.3)$$

where $s_f(y_t \mid x_{t-1}, \tilde{\theta}_n) = \frac{\partial}{\partial\theta}\ln f(y_t \mid x_{t-1}, \tilde{\theta}_n)$ denotes the quasi-score function. The dimension of the auxiliary model, l_θ, is selected by following an upward model expansion path, using the Schwarz's Bayesian information criterion (BIC) $BIC = s_n(\tilde{\theta}) + (l_\theta / 2n)\ln(n)$, where $s_n(\tilde{\theta}) = -L_n(\tilde{\theta}, \{y_t\}_{t=1}^{n})$ is the negative maximized objective function. Implied by standard QML theory, even if the auxiliary model is misspecified, under suitable regularity, $\tilde{\theta}_n \xrightarrow{p} \theta_0$, where the limiting value, θ_0, is denoted the quasi-true value of θ.

The projection step provides a summary of the data, which will be used as the score generator for the next step of estimation. Gallant and Long (1997) show that a judicious selection of the auxiliary model, ensuring that it approximates the salient features of the observed data, will result in full asymptotic efficiency. Effectively, as the score generator approaches the true conditional density, the estimated covariance matrix for the structural parameter approaches that of maximum likelihood. This result embodies one of the main advantages of EMM. It prescribes a systematic approach to the derivation of efficient moment conditions for estimation in a general parametric setting.

1.2.2 Estimation Step

In the estimation step estimates of the parameters of a candidate structural parameter are obtained from a GMM-type estimation procedure using the fitted scores

from the SNP model as the moment conditions. To do this, for a specific structural model represented by $P(y_t \mid x_{t-1}, \rho)$ with a given parameter vector ρ, a simulated series $\{\hat{y}_t\}_{t=1}^{N}$ is generated. Identification requires that the dimension of the quasi-score (the length of θ), l_θ, exceeds that of the structural parameter vector, l_ρ. An average over a long simulation from the true structural model, reevaluated at the fixed QML estimate,

$$m_N(\rho, \tilde{\theta}_n) = \frac{1}{N}\sum_{n=1}^{N} \frac{\partial}{\partial \theta} \ln f(\hat{y}_t(\rho) \mid \hat{x}_{t-1}(\rho), \tilde{\theta}_n) \tag{1.4}$$

would satisfy $m(\rho_0, \tilde{\theta}_n) = 0$. In the usual case in which $l_\theta > l_\rho$, the structural parameters ρ are estimated by minimizing the EMM objective function

$$\hat{\rho}_n = \arg\min_{\rho} \left[m_N(\rho, \tilde{\theta}_n)' \tilde{I}_n^{-1} m_N(\rho, \tilde{\theta}_n) \right] \tag{1.5}$$

where $\tilde{I}_n$ denotes a consistent estimator of the asymptotic covariance matrix sample quasi-score vector. The estimate $\tilde{I}_n$ is obtained directly from the first step which avoids the need for computation of the weighting matrix during the second GMM-based estimation step. In addition, if the auxiliary model is expanded to the point where it accommodates all main systematic features of the data, likelihood theory implies that the quasi-scores constitute a (near) martingale difference sequence, and a convenient estimator of the quasi-information matrix is obtained from the outer product of the score:

$$\tilde{I}_n = \frac{1}{n}\sum_{t=1}^{n} \left[\frac{\partial}{\partial \theta} \ln f(y_t \mid x_{t-1}(\rho), \tilde{\theta}_n) \right] \left[\frac{\partial}{\partial \theta} \ln f(y_t \mid x_{t-1}(\rho), \tilde{\theta}_n) \right]' \tag{1.6}$$

Gallant and Tauchen (1996) show that, under suitable regularity conditions, the EMM estimator $\hat{\rho}_n$ is almost surely consistent and asymptotically normal. Moreover, the asymptotic variance-covariance matrix may be estimated consistently by

$$\mathrm{cov}(\hat{\rho}_n) = \frac{1}{n}\left[\frac{\partial m_N(\hat{\rho}_n,\tilde{\theta}_n)'}{\partial \rho}\tilde{I}_n^{-1}\frac{\partial m_N(\hat{\rho}_n,\tilde{\theta}_n)'}{\partial \rho'}\right]^{-1} \tag{1.7}$$

The usual GMM test of over-identifying restrictions may be used to test model adequacy. If the structural model is correctly specified, then the normalized EMM objective function satisfies

$$n m_N(\hat{\rho}_n,\tilde{\theta}_n)'\tilde{I}_n^{-1} m_N(\hat{\rho}_n,\tilde{\theta}_n) \square \chi^2(l_\theta - l_\rho) \tag{1.8}$$

If the overidentification test rejects an underlying structural model, the individual elements of the score vector may provide useful information regarding the dimensions in which the structural model fails to accommodate the data. These model diagnostics are based on the standard *t*-statistics of the individual elements of the score vector, $m_N(\hat{\rho}_n,\tilde{\theta}_n)$. Obtained by normalizing the score vector by its standard error, these *t*-statistics can be interpreted much as normalized regression residuals. Thus, large *t*-ratios reveal those characteristics that are not well approximated. Subject to the same risk as the interpretation of regression residual, the *t*-ratios are usually biased downward, and therefore conservative. Nonetheless, as with regression residuals, inspecting normalized elements of $m(\hat{\rho}_n,\tilde{\theta}_n)$ is usually the most informative diagnostic available.

Another advantage of using EMM estimation is the ability to rank non-nested structural models. Notice that the weight matrix in GMM used in constructing the specification test is identical across different model specifications. Consequently, the *p*-

value based on the overidentification test can be directly compared across different structural models to identify the best structural model. It is well recognized in the literature that tests for the presence of regime shifts against an alternative require nonstandard approaches. Our approach of comparing all the considered models to a common nonparametric density allows us to rank order all the considered models according to the p-values implied by the EMM criterion function.

1.3. SV Models for the Short Rate

In this section we discuss a series of models and extensions to explain short-term interest rate dynamics. The first type of model is the generalized Gaussian diffusion model that is commonly used in building term structure models. To incorporate additional factors, we extend the one-factor diffusion model to the two-factor and three-factor SV models that has been proven to be more successful than the ARCH/GARCH model in modeling the dynamics of the second moment of many financial time series. The second type of model is the non-Gaussian diffusion model with Stable distributed innovations, which has recently become popular in the empirical finance literature. The third type of model allows for Markov RS behavior in the specification of the volatility dynamics, with the flexibility of simultaneously mixing the RS effect with the SV effect and the level effect of volatility. The Gaussian and non-Gaussian diffusion models are continuous-time models and the RS models are discrete-time models.

1.3.1. Gaussian Diffusion Models

A. One-Factor Gaussian Diffusion Model

Firstly, we consider the generalized diffusion model, presented by Chan et al. (1992), in which the instantaneous change in the short rate can been characterized as a stochastic differential equation (SDE) given by

$$dr_t = (\phi_0 - \phi_1 r_t)dt + \sigma r_t^{\gamma} dW_1 = k_r(\mu_r - r_t)dt + \sigma r_t^{\gamma} dW_1 \qquad (1.9)$$

where $\{r_t\}$ is the short rate at time t, and dW_1 is a standard Wiener process. We call (1.9) the CKLS model. The key characteristic of the dynamics is that the conditional mean and variance of changes in the short rate depend on the level of the rate. Specifically, in this model, r_t mean-reverts towards the long-run level μ_r, with the speed of the reversion measured by k_r, and γ captures the so-called "level effect" in which of the level of rates influences the conditional volatility. By allowing γ to be estimated freely, many well known models can be nested with appropriated parameter restrictions within this generalized model.[5]

To empirically calibrate the general SDE (1.9), Chan et al (1992) employed the following discretization approximation

$$\Delta r_t = \phi_0 - \phi_1 r_t + \sigma r_t^{\gamma} z_{t+1} = k_r(\mu_r - r_t) + \sigma r_t^{\gamma} z_{t-1} \qquad (1.10)$$

and estimated the model parameters with the generalized methods of moments (GMM) estimation technique of Hansen (1982). Using monthly data from 1964-1989, they found that the short rate was mean reverting, and reported a point estimate of 1.4999 for the level effect parameter γ which implies the volatility of short-term interest rates is

[5] See Chan et al (1992).

explosive. With similar data, Smith (2002) estimated the CKLS model using a quasi-maximum likelihood methodology and reported a similar level effect estimate of 1.4515. In order to obtain the maximum likelihood estimates and guarantee a compatible comparison with the SV and RS models, Smith used a two step procedure for estimating the models. In the first step, he used the ordinary least square (OLS) to obtain an estimate of the mean reversion parameters ϕ_0 and ϕ_1. In the second step, he formed the fitted residuals $\Delta r_t = \phi_0 - \phi_1 r_{t-1} + e_t$, and then estimated the remaining parameters from the transformation of the log of the squared residual. This estimation procedure is required to build up the likelihood functions for the SV and RS models.

Although these findings are instructive for understanding the short-term rate dynamics, they are not entirely satisfactory. First, Monte Carlo studies have questioned the efficiency of using GMM estimation in sense of the choice of the moment conditions and its finite sample performance. The two-step estimation procedure used in Smith (2002) suffers from the loss of the estimation efficiency as well. Lastly, evidence has been shown that the estimated parameters of the CKLS model are sensitive to the data frequency. In particular, the level effect parameter estimate from monthly data could be spuriously high and unstable; using more frequently sampled data leads to different results. In addition, as pointed out by Andersen and Lund (1997), the internal dynamics proposed in the discrete-time models, at estimated parameter values, are excessively erratic. This severely limits their usefulness for numerical or simulation-based estimation procedures. To avoid the previously mentioned difficulties in estimating models of the short rate, in this paper we rely on the EMM estimation

using weekly data and estimate the continuous-time CKLS model directly rather than using a discretization approximation.

B. Two-Factor SV Model

We consider the following CKLS model extended to have stochastic volatility in the spirit of Taylor (1986, 1994):

$$\begin{cases} dr_t = (\phi_0 - \phi_1 r_t)dt + r_t^{\gamma}\sigma_t dz_t = k_r(\mu_r - r_t)dt + r_t^{\gamma}\sigma_t dW_1 \\ d\log(\sigma_t^2) = (\omega_0 + \omega_1 \log(\sigma_t^2))dt + \xi dW_2 \end{cases} \tag{1.11}$$

where dW_1 and dW_2 are mutually independent i.i.d. Wiener processes. For these dynamics, the log-volatility of short rate series is assumed to follow a mean reverting process as well as the series itself. Also, the conditional volatility is subject to random shocks, and the sensitivity to these shocks is measured by the parameter $\xi > 0$.

Maximum likelihood estimation is generally not feasible for estimating the SV models due to the presence of an unobserved volatility. One procedure available is the quasi-maximum likelihood procedure of Harvey, Ruiz, and Shephard (1994). This approach uses a transformation on the log of the squared residual in order to write the system in state-space form, and then applies the Kalman filter to recursively build up the likelihood function. Smith (2002) followed this two-step estimation procedure and reported an estimate of 1.44 for the level effect parameter using monthly data.

Andersen and Lund (hereafter AL) (1997) estimated (1.11) directly using the EMM estimation technique with a SNP auxiliary model that employs a Level-EGARCH leading term. Using weekly data over the 1954-1995 sample periods, they found the level effect parameter to be close to 0.5. While their model was rejected by the data at

the 5 percent significant level, the incorporation of the unobservable volatility factor was shown to greatly enhance the model's ability to fit the data and the implied process was much less erratic than the process implied by the CKLS estimates[6]. Following AL, we estimate (1.11) using the EMM estimation with a longer span of weekly data.

C. Three-Factor SV Model

We consider the following continuous-time three-factor SV model for the short rate:

$$\begin{cases} dr_t = (\phi_0 - \phi_1 r_t)dt + \sigma_t r_t^{\gamma} dz_t = k_r(\mu_{r,t} - r_t)dt + \sigma_t r_t^{\gamma} dW_1 \\ d\log(\sigma_t^2) = (\omega_0 + \omega_1 \log(\sigma_t^2))dt + \xi dW_2 \\ d\mu_t = (\upsilon_0 + \upsilon_1 \mu_t)dt + \zeta dW_3 \end{cases} \tag{1.12}$$

where dW_1, dW_2 and dW_3 are mutually independent i.i.d. Wiener processes. In (1.12), the log-volatility of short rate series and the long-run mean are assumed to follow mean-reverting process. The sensitivity of shocks to the log-volatility and to the long-run mean are measured by the non-negative parameters ξ and ζ, respectively.

The model (1.12) is an extension of the two-factor SV model (1.11) suggested by the AL. The introduction of a third factor associated with the reverting mean level may improve the data fitting through accommodating the time-varying drift behavior over the sample period. According to AL, time variation in the reverting mean could be interpreted as variation in an underlying inflation rate.

---

[6] A number of other estimation procedures have been implemented for the two-factor SV model, including the Bayesian technique of Jacquier, Polson, and Rossi (1994), the maximum likelihood procedure of Fridman and Harris (1998), and the maximum likelihood Monte Carlo method of Sndmann and Koopman (1998).

The three-factor model (1.12) is a particular form of a general class of affine multifactor models. Dai and Singleton (2002) discussed the general issues for the identification and admissibility conditions of affine diffusion models, which are characterized by linearity of the drift and variance functions. The investigation of other types of three-factor SV models or general affine diffusion models is left for further research.

1.3.2. Non-Gaussian Diffusion Model

The modern asset pricing theory and, more specific, the option pricing theory have been firmly built upon the Gaussian diffusion framework based on the beliefs that the financial data tends to become more Gaussian over longer timescales. The popularity of the SV approach is partially due to its consistency with the Gaussian assumption making possible an appropriate generalization of the Black-Scholes option pricing framework. However, empirically studies have shown that financial returns exhibit features that are incompatible with the assumption of Gaussian data. The leptokurtosis implied by the Gaussian diffusion and SV models tend to be far less than the sample kurtosis observed from many financial series, although the implied time-varying and persistent volatilities are consistent with the data.

One generalization developed to explain the observed leptokurtosis and skewness is the jump-diffusion model originally proposed by Merton (1976). This model consists of two parts: a continuous part modeled by a geometric Brownian motion, and a jump part with the logarithm of the jump sizes having a double exponential distribution and the jump times corresponding to the event times of a Poisson process. General

properties of jump-diffusion models with independent identically distributed jump sizes have been extensively studied; for an excellent survey, see Duffie (2000).

Another generalization is to consider diffusion models assuming non-Gaussian distributions to capture the departures from the Gaussian diffusion model. Following this direction, a variety of non-Gaussian distributions has been considered in discrete-time models.[7] However, these models suffer from the lack of "stability"; i.e., the distribution of the increments do not depend on the time intervals, which is a desirable property for asset returns particularly in the context of portfolio analysis and risk management as stressed by Mandelbrot (1963). In fact, the stable law[8] is the only possible weak limit of properly normalized sums of i.i.d. random variables and only for stable distributed returns do we have the property that linear combinations of different return series follow again a stable distribution.

Motivated by the nice properties of the stable law and stability under-addition, we consider the following continuous-time non-Gaussian CKLS model with stable Lèvy increments

$$dr_t = (\phi_0 - \phi_1 r_t)dt + \sigma r_t^{\gamma} dW_1 = k_r(\mu_r - r_t)dt + \sigma r_t^{\gamma} dL_t \qquad (1.13)$$

The key characteristics of this model are essentially the same as those of Gaussian CKLS model (mean-reverting process for the drift dynamics and the incorporated level effect for the variance dynamics of the short rate), except that L_t is a stable Lèvy process.

---

[7] See a review of, among others, Peters (2001) and Verhoven and McAleer (2003).

[8] The sum of a number of random variables with power-law tail distributions having infinite variance will tend to a stable Lèvy distribution as the number of variables grows, also referred to as the generalized central theorem.

Stable Lèvy processes are stochastic processes with independent and stationary increments. A stochastic process L_t is a stable Lèvy process if and only if: (1) it has independent increments; that is, for $0 < a < b < c < d$, $L_d - L_c$ and $L_b - L_a$ are independent; (2) it has stationary increments; that is, the distribution of $L_{t+s} - L_t$ does not depend on t; (3) it is stochastically continuous (4) with probability one it has right-continuous paths with finite left-limits; and (5) $L_0 = 0$ almost surely. The Brownian motion is a special example of Lèvy processes, one which is with stationary, independent increments having a Gaussian distribution; here we consider the standard stable process, which is a Lèvy process with stationary, independent increments having a standard stable distribution, $S(\alpha, \beta, 0, 1)$. A stable distribution is characterized by four parameters: $(\alpha, \beta, c, \delta)$. The exponent, α, confined to the interval $0 < \alpha \leq 2$, is known as the shape variable, which influences the total probability contained in the extreme tails, or the shape of the distribution. The smaller the value of α, the thicker the tails of the distribution. In particular when $\alpha = 2$ we get the normal distribution. The parameter β measures asymmetry of the distribution. If $\beta = 0$, the distribution is symmetric about the location parameter δ; if $\beta = 1$, the distribution is totally skewed to the right and similarly it is totally to the left when $\beta = -1$. The scale parameter c narrows or broadens the distribution about δ in proportion to c. A standard stable distribution has $\delta = 0$ and $c = 1$. Note a stable distribution given by $(2, 0, c, \delta)$ is exactly a normal distribution with mean δ and variance $2c^2$.

There is not much published literature on empirical volatility modeling using stable distributions. The non-existence of moments of second or higher order is a major

drawback of the use of the stable distribution from an empirical point of view. Also, with the exception of a few cases, the probability density function is not known in closed form; therefore, one has to use their characteristic functions instead. On the other hand, one can use stable distributions to save the CLT argument, based on which a similar asset pricing framework to the current Gaussian one could be established; it also can easily accommodate heavy tails and skewness of financial series, which is a much desired property in empirical finance. For these reasons, the use of stable processes has recently become substantially more popular in the modeling of stochastic volatility (Liu and Brosen (1995)), portfolio theory (Olotarev (1986), Mittnik and Rachev (1991), Cheng and Rachev (1995)), asset pricing theory (Connor (1984), Gamrowski and Rachev (1994,1995)), option pricing (Rachev and Samorodnitsky (1993), Janicki and Weron (1994), Bouleau and Lepingle (1994), Matacz (2004)), and other financial phenomena.[9]

In our estimation of (1.13) using EMM, we fix the characteristic parameters α and β of the stable distribution and freely estimate the remaining parameters. The choices for α and β are *ad hoc* and it would be desirable to estimate these parameters freely.[10]

1.3.3. RS Models

The diffusion models discussed in the previous subsections are single-regime models in that they have a single structure for the conditional mean and variance. For

[9] See Marinelli and Rachev (2002).

[10] Garcia, Renault and Veredas (2004) discussed the estimation of the parameters of a Stable distribution using the indirect inference methods relative to other prevalent methods based on the characteristic function and the empirical quantiles of the Stable distribution.

example, the CKLS model for the short rate is assumed to be mean reverting to the same long-run mean, with the same speed of reversion and the same level effect throughout the sample. A more flexible extension is to relax the assumption of a single regime in favor of a two-state Markov RS specification. Many authors have proposed RS models for fitting the dynamics of the short-term interest rate (see, Hamilton (1998), Garcia and Perron (1996), Gray (1996) and Ang and Bakeart (2001), Liechty and Roberts (2001)), for the impact on the entire yield curve using dynamic term structure models (see, Naik and Lee (1997), Boudoukh et al. (1999), Evans (2001) and Bansal and Zhou (2003), Dai, Singleton and Yang (2004)), and for the bond pricing in the RS context (see, Landén (2000) and Wu and Zeng (2003)). While many theoretical and empirical works show strong evidence for regime switching in interest rates, the specification issue of the RS model for the conditional mean and variance dynamics of the interest rates has not been extensively explored in the literature. Considering that our interest in this paper is to model the volatility dynamics for the short rate, we assume a simple specification in which the conditional mean parameters are regime independent. Furthermore, in our specification of RS models we use the discrete-time approximation to the continuous-time diffusion used in CKLS (1992), which is consistent with the rationale that large regime switching behavior only occurs infrequently over time. EMM estimation enables us to compare the RS models with the continuous-time models based on the EMM objective function p-value.

Given the assumption of the single-regime conditional mean dynamics, we propose four RS models to describe the volatility dynamics. The first model is a

simplified regime switching-in-volatility model (RS-in-σ model hereafter) based on a discretized OU process, given by

$$\Delta r_t = \phi_0 - \phi_1 r_{t-1} + \sigma_i z_t \qquad i = 1,2 \qquad (1.14)$$

This model assumes the same speed of mean reversion ϕ_1 to a common long-run mean (ϕ_0 / ϕ_1), but allows different shocks within each regime to accommodate time-varying volatility. The switching states are governed by a first-order Markov process. The time invariant transition probabilities from regime j to regime i are defined as $p_{ij} = \Pr(S_t = j \mid S_{t-1} = i)$ with the restriction $\sum_{i=1}^{2} p_{ij} = 1$. For the case of two states, the matrix of transition probabilities is given by [11]

$$P = \begin{pmatrix} P_1 & 1-P_1 \\ 1-P_2 & P_2 \end{pmatrix} \qquad (1.15)$$

Due to the success of the two-factor SV model over the one-factor diffusion model, the second RS model we consider is an extension of the $RS-in-\sigma$ model where the conditional variance is driven by a SV process. This model ($RS-in-\sigma+SV$ model hereafter) is given by

$$\begin{cases} \Delta r_t = \phi_0 - \phi_1 r_{t-1} + \sigma_t z_t \\ \Delta \log(\sigma_t^2) = \omega_{0,i} + \omega_1 \log(\sigma_{t-1}^2) + \xi u_t \end{cases} \quad \text{with } i = 1,2 \quad (1.16)$$

[11] For future research, the RS models can be generalized to have a greater number of states or the regime switching probabilities can be made a function of the level of interest rates. The latter case allows for the possibility that a switch to the high-volatility regime may be more likely when interest rates are high according to Gray (1996).

The conditional variance of (1.16) has a regime independent random shock but regime-dependent reverting mean. Thus, the $RS-in-\sigma+SV$ model nests the simple OU process, OU-SV process, and $RS-in-\sigma$ model as special cases.

Different from the above two RS models built on the OU process, the following two RS models are based on the generalized CKLS process. Incorporating both the RS-in-volatility effect and the level effect, the third RS model is called the $RS-in-\sigma+Level$ model and is given by

$$\Delta r_t = \phi_0 - \phi_1 r_{t-1} + \sigma_i r_{t-1}^{\gamma} z_t \qquad \text{with } i=1,2 \quad (1.17)$$

The $RS-in-\sigma+Level$ model incorporates the sensitivity of volatility to the current level of short rate, measured by γ, to accommodate additional time-varying behavior and conditional heteroskedasticity, although the level effect parameter is kept the same across the regimes.

The fourth RS model is an extension of the $RS-in-\sigma+Level$ model, which we call the $RS-in-\sigma+Level+SV$ model, is given by

$$\begin{cases} \Delta r_t = \phi_0 - \phi_1 r_{t-1} + \sigma_t r_{t-1}^{\gamma} z_t \\ \Delta \log(\sigma_t^2) = \omega_{0,i} + \omega_1 \log(\sigma_{t-1}^2) + \xi u_t \end{cases} \qquad \text{with } i=1,2 \quad (1.18)$$

In addition to the characteristics of the $RS-in-\sigma+Level$ model, the conditional log-volatility process is driven by a SV process, with regime-dependent mean reversion ($\omega_{0,i}$) but regime-independent random shocks. The $RS-in-\sigma+Level+SV$ model nests as special cases discrete-time versions of the CKLS model, the two-factor SV model, and the three other RS models.

The first two RS models based on a simple OU process are motivated by the work of Gray (1996). He used a generalized RS framework where all conditional mean parameters (ϕ_0 and ϕ_1) and conditional variance parameters (σ) are allowed to switch across the two regimes. He considered a different extension of the $RS-in-\sigma$ model where the conditional variance is driven by a GARCH process rather than an SV process. Using weekly data on the 30-day T-bill rate, he argued that both the RS effect and the GARCH effect are important to adequately fit the data. He also constructed a likelihood ratio test to compare his $RS-in-\sigma$ model with his $RS-in-\sigma+GARCH$ model.

The last two RS models, based on the CKLS model, are motivated by Smith (2002). He employed a two-step procedure in order to overcome the difficulty of estimating the RS model using the quasi-maximum likelihood approach of Harvey, Ruiz, and Shephard (1994). Smith showed that the level effect parameter is spuriously high in the single-regime models, and is reduced to around unity in his RS models. He also argued that either the SV effect or the RS effect, but not both, are needed for describing the data accurately. We note that So, Lam, and Li (1998) developed a similar model as our $RS-in-\sigma+Level+SV$ model and estimated it using the Bayesian technique of Jacquier, Polson, and Rossi (1994).

For RS models, EMM estimation has advantages over the QML Kalman filter procedure and other estimation techniques. With EMM, we can estimate all the unknown parameters simultaneously to ensure that no important information has been lost in the process, which cannot be guaranteed by the two-step procedure of Smith (2002). Another problem that may relate to the efficiency loss of Smith (2002) is that

the simulated conditional volatility process based on his parameter estimates (especially the positive volatility persistency parameter) is a highly explosive process. In addition, the usual test statistics cannot be applied to test the existence of the second regime since parameters associated with the second state are unidentified under the null of one regime. Most of the past works obtained the evidence for the existence of the additional regime from the enormous increase in the likelihood value when moving from a single-regime model to a two-regime model or carefully applied the LRT to compare the regime-switching models. With EMM, all the comparable models could be easily ranked according to the simple measurement of the p-values implied by the EMM criterion function. After the one-to-one model comparison, we expect to have a systematic answer for questions such as (1) whether the simple $RS-in-\sigma$ model could mimic the performance of complicated non-Gaussian diffusion models, (2) whether $RS-in-\sigma+Level$ model could save the efforts of adding one stochastic factor as implied by SV models, and (3) which effect or effects among the three, the level effect, the SV effect, and the RS effect, are needed to adequately fit the data of US short rates.

1.4. Data

Our empirical work uses weekly (Wednesday) observations of the annualized yield on the 3-month U.S. T-bill over the period January 1954 to September 2004, forming 2648 observations. The data was constructed from a daily series available from the Federal Reserve Bank, where the rates are calculated as unweighted averages of closing bid rates quoted by at lease five dealers in the secondary market, and the rates

are posted on a bank discount basis, but converted into continuously compounded yields prior to analysis. We analyze weekly rates over daily rates to avoid missing data, possible holiday and weekday effects, and other potential problems associated with market microstructure effects. Wednesday data are used because of the least number of missing observations for this weekday. When a Wednesday rate is missing, we use the Tuesday rate; when a Tuesday rate is missing, use the Thursday rate. The data preparation procedure follows Andersen and Long (1997).

The raw data plotted in Figure 1.2, and descriptive statistics are given in Table 1.2. The basic stylized facts concerning the short-rate are: near nonstationary behavior (slow mean reversion), large changes and small changes are clustered together (ARCH effect), the volatility of rates increases with the level of rates (level effect), and positive skewness and excess kurtosis[12] (non Gaussian distribution). The non Gaussian behavior of the short rate is clearly shown in the qq-plot in Figure 1.4 and in the statistics summary in Table 1.2, and the slow mean reversion and ARCH effect are illustrated in the autocorrelation plots in Figure 1.3.

The data period of our sample, 1954 to 2004, represents the longest weekly set of observations on the 3-month T-bill rate, which is important for evaluating models that purport to explain mean and volatility dynamics. Also, our sample contains seven major recessions and six major expansions, which provides economic motivation for incorporating regime shifts into the models. Some important events that may cause strong shifts in the behavior of interest rates dynamics include: the Vietnam War from 1961 to 1975, the simultaneous occurrence of recession and inflation in the early 1970s,

[12] Kurtosis of the Gaussian distribution is three; excess kurtosis for a non-Gaussian distribution is the different between its kurtosis and three.

the 1973 energy crisis due to the onset of an oil embargo by OPEC until 1975, the "Monetary experiment" conducted by the Federal Reserve during 1979-82 when its policy shifted away from targeting federal fund rate, the largest stock market crash on October 19, 1987, the Gulf War which started in August 1990, and the longest peacetime economic expansion in U.S. history beginning in March 1991.[13] The period from 1996 to 2004, which was not covered by many previous analyses of the short rate, poses an especially tough challenge for standard asset pricing models. This period started with an unprecedented period of long economic growth and a bull stock market run, which was interrupted by the September 11, 2001 terrorist attack, and was followed by a downturn of the stock market, and finally ended with the "War on Terrorism" campaign with the invasion of Iraq on March 2003.[14]

1.5. Empirical Results

1.5.1. Estimation of the SNP Auxiliary Model

The first step in EMM estimation is to project the observed data onto an auxiliary model that captures all of the relevant characteristics of the data. We use the semi-nonparametric (SNP) conditional density model described in Gallant and Tauchen (2001) as our auxiliary model. The selection of an appropriate auxiliary model is essential for the success of EMM estimation, especially for interest rate data as stressed by Andersen and Lund (1997) and Gallant and Tauchen (2004). The empirical

[13] See Choi (2004).
[14] See Bansal, Tauchen and Zhou (2003).

literature on EMM estimation of the short-rate, however, has not explored the relevance of this issue in a systematic manner.[15]

We follow Gallant and Tauchen (2001) and use a specific-to-general model selection procedure based on minimizing a Bayesian information criterion (BIC). In particular, the SNP tuning parameters $(L_u, L_g, L_r, L_p, K_z, I_z = 0, K_x, I_x = 0)$ are selected by moving upward along a model expansion path where small values of BIC are preferred. The expansion paths we follow are illustrated in Table 1.3. First, the autoregressive order L_u is determined. The expansion path with ARCH leading terms is to expand L_r, then to expand K_z, and finally expand K_x. For GARCH leading terms, the strategy is to put $L_r = L_g = 1$ first, then expand K_z and K_x. The expansion paths we follow are not exhaustive across models and it sometimes happens that the best set of the tuning parameters lies elsewhere within the expansion path. Therefore, we also explore some other paths which slightly deviate from the ones specified in Table 1.3.

The best fitting SNP models for the 3-month T-bill rate in terms of BIC, characterized by the set of tuning parameters, $(L_u, L_g, L_r, L_p, K_z, I_z = 0, K_x, I_x = 0)$, are reported in Table 1.4. Following the upward BIC protocol and exploring beyond the expansion path a bit, the preferred auxiliary model is the SNP 11117000 model. The SNP 11117000 model is a GARCH (1,1) model with a nonparametric error density represented as a seven-degree Hermite polynomial expansion of the normal density where the Hermite coefficients are state independent. The model is similar to the semi-parametric GARCH of Engle and Gonzalez-Rivera (1991). Table 1.5 gives the

[15] See Brandt and Chapman (2002).

parameter estimates. The estimated AR coefficient is 0.999 which implies a very slow mean reversion and near nonstationary behavior. The sum of the ARCH and GARCH terms implies highly persistent conditional volatility.[16] The large positive coefficient on the 4^{th} order Hermite term and the positive coefficient on the seventh order Hermite term capture the fat tails and positive skewness in the demeaned short-rate series. Our preferred SNP model for the short rate is similar to the SNP models used by other authors as shown in Table 1.6.

As stressed by Gallant and Tauchen (2001), if the fitted SNP model is to be used as the score generator in conjunction with EMM it is important to check the dynamic stability of the model. For complicated SNP models, a simple way to check dynamic stability is to generate long simulations from the fitted model and observe if these simulations become explosive. For non-explosive models, the simulations should capture all the salient feathers of the observed data. The simulated series based on the fitted SNP models are shown in Figure 1.5. From the plots, it can be observed that the 11117000 SNP model mimics the observed data fairly well, although it produces simulations with negative interest rates. The simulation from the 11118000 SNP model is also plotted in Figure 1.5, and it appears mildly explosive.

Residual diagnostic checks on the fitted model are conducted to verify that it is adequate and appropriate. Panel (A) in Figure 1.6 gives the estimated conditional volatilities from the 11117000 model, and these capture the observed volatility patterns in the observed data. Panel (B) shows the estimated conditional density, which is more peaked in the center with heavy tails relative to the Gaussian distribution. The qq-plots

[16] Because of the absolute value formulation in the GARCH specification, the sum of the ARCH and GARCH coefficients do not have to be less than one for the model to be stationary.

for the simulated series from the fitted SNP 11117000 model and its first order change are shown in Panel (C) and Panel (D), both of which capture the patterns of the real data series. The standardized residuals, shown in Panel (E), mostly resemble a white noise process. However, there are some large outliers present. The autocorrelation plots of the residuals and squared residuals in Panel (F) reveal no significant autocorrelation and indicate that the fitted SNP model adequately captures the conditional dynamics in the mean and volatility.

1.5.2. EMM Estimation Results

In this subsection we report the EMM estimation results for a number of structural models for interest rates described in Section 3.2. The single regime structural models we consider are: the one-factor CKLS model with Gaussian errors (CKLS-N), the two and three-factor SV model (SV2, SV3), the non-Gaussian stable diffusion model with shape variable α and skewness variable β (CKLS-S(α, β)). The Markov regime switching (RS) models we consider are: the RS-in-σ and RS-in-$\sigma + SV$ model based on a simple OU process, and the RS-in-$\sigma + Level$ and RS-in-$\sigma + Level + SV$ model based on the generalized CKLS model.

The EMM estimation procedure requires the simulation of a long sample from the underlying structural models. For both the discrete-time and continuous-time models, the EMM objective function is formed using a default simulation size of 75,000, where we have discarded the first 5,000 observations. Restarts of the optimizer at random perturbation of the initial value values are employed for EMM to avoid local optima. For continuous-time diffusion models, the simulations are generated by the Euler

scheme using 25 subintervals per week.[17] Tables 1.7-1.9 contain the results for short rate from estimating each model outlined above.[18] In the following sections, we present in detail the one-by-one model performance and comparison.

A. Gaussian Diffusion Models

Table 1.7a reports the EMM estimation results for the Gaussian diffusion models: the one-factor Gaussian CKLS model, the two-factor SV model, and the three-factor SV model. The small p-value based on the χ^2 distribution associated with the EMM objective function values, leads to a strong rejection of the one-factor Gaussian CKLS model. On the other hand, the two-factor SV model and the three-factor SV model are not rejected at the 10% level; the former is in contrast to what Andersen and Lund (AL) (1997) found. Our results indicate that the introduction of an additional stochastic volatility factor is important for explaining observed interest rate behavior.

Our estimation results suggest the following insights about the dynamics of the short rate. The signs of all the parameter estimates for the mean dynamics are consistent with the GMM estimates of the one factor CKLS model reported in Chan et al (1992) based on monthly data. All of the models indicate that short rates revert ($-\phi_1 < 0$) to a positive long-run mean ($\phi_0 / \phi_1 > 0$), with a very slow rate of mean reversion.[19] Based on our estimates of the two-factor SV model, the implied estimated measure of log-volatility persistence, $\exp(-\omega_1 / 52)$, is about 0.9893 at the weekly level,

[17] Further details regarding the implementation refer to the appendix in AL (1997).

[18] The estimation is conducted using the S-PLUS implementation of Gallant and Tauchen's EMM FORTRAN code available in S+FinMetrics 2.0 and described in Zivot and Wang (2006).

[19] Notice that the reason that the long-run reverting mean for the short rates implied by the estimation of our one-factor CKLS models differs substantially from the GMM estimation is due to the fact that we use percentage interest rates for the analysis, rather than decimal interest rates as Chan et al (1992).

and the discrete-time autoregressive coefficient in the mean dynamics, $\exp(-\phi_1/52)$, is about 0.994. These estimates are comparable with those reported in AL. Moreover, we find that the conditional volatility of rates is sensitive to the level of the rates; that is, the elasticity of volatility measured by γ is significantly in excess of zero.

The incorporation of data after 1989 in the estimation changes the implied dynamics of the short rate substantially in many aspects from previous estimates. For example, our two-factor SV model implies a lower long-run mean, measured by ϕ_0/ϕ_1, of 2.85% and a faster speed of mean reversion than the results found by AL using a similar model. This difference may be partially explained by the fact that the Federal Reserve started to decrease the Federal Funds rate dramatically after 2001 in order to boost the economy after the "9/11" recession. More striking is the change in the estimate of the level effect, measured by γ. The level effect estimate in our one-factor CKLS model is 0.3 which is substantially lower than the GMM estimate, $\gamma = 1.49$, reported in CKLS (1992). AL showed that the level effect is weakened if a second volatility factor is incorporated. Our results show that the evidence for a strong level effect is significantly weakened without an additional SV factor. The estimate of the level effect in our two-factor SV model is 0.67 which is a bit larger than what AL found in their SV2 model.

Our SV3 model involves the introduction of a third factor associated with the mean level as suggested by AL who suspected that a time-varying long-run reverting mean as well as a time-varying conditional volatility is needed to accommodate the

data.[20] We are not surprised to see that the three-factor SV model is favored over the two-factor SV model and the result improves significantly by adding this mean related factor. Implied from the SV3 model, the short rate process is reverting to a time-varying unconditional mean, which itself is also a mean reverting process with reverting trend measured by υ_0/υ_1, of 2.50%. This estimate is close to the reverting mean implied from the previous SV2 model. Moreover, the corresponding level effect estimate is lowered to round 1/2, which is slightly smaller than that of the two-factor SV model.

Part of Table 1.8a displays the diagnostics for all Gaussian diffusion models, based on the informative standard *t*-ratios of the individual elements of the score vector. These *t*-statistics can be interpreted conveniently as normalized regression residuals. Therefore, large *t*-ratios reveal those characteristics that are not well approximated. It appears that the one-factor CKLS model encounters difficulties to accommodate the scores; the large *t*-ratios on individual score elements associated with the second to sixth Hermite polynomial elements show that the it fails to capture certain aspects of volatility clustering that exists in the data as summarized by the 11117000 auxiliary model. On the other hand, for the accepted two-factor and three-factor SV models, all adjusted *t*-ratios are well below 2.0.

B. Non-Gaussian Diffusion Model

Table 1.7b reports the EMM estimation results for the non-Gaussian stable diffusion models with fixed combinations of the shape and skewness parameters. The

[20] See also Gallant and Tauchen (2002), in which they proposed a two-factor SV model with a mean factor for the Microsoft stock returns.

small p-values for the EMM objective function lead to rejections for all the non-Gaussian diffusion models at the 5 percent significance level. The score diagnostics provided in Table 1.8 provide some explanations for the failure of these models. For example, the CKLS model with stable $(\alpha = 1.95, 0)$ errors fails to capture certain aspects of volatility clustering associated with the third and fifth order Hermite polynomial elements. The best fitting model, with a p-value of 0.045, is the CKLS-S$(\alpha = 1.9, \beta = 0.1)$ model. All of the score t-ratios for this model are smaller than two.

The parameter estimates from the best fitting CKLS-S$(\alpha = 1.9, \beta = 0.1)$ model are similar to the one-factor Gaussian diffusion model, implying a strong mean reversion in the short-term rates and a slightly larger level effect that is less than unity. Overall, the stable diffusion models with $\alpha \geq 1.9$ have higher χ^2 values than the Gaussian model which shows that allowing for heavier tails for the innovation density improves the model fit. In addition, the best fitting CKLS-Stable model has a positive skewness parameter $\beta = 0.1$, implying that fat-tailed *and* positive skewed innovations are important for explaining the data.

The one-factor CKLS-S$(\alpha = 1.9, \beta = 0.1)$ model that allows for fat tails and positive skewness can accommodate many of the complex features of the interest rate series. This model can accommodate outliers much more easily than the Gaussian model, and its fit is similar to the three-factor continuous-time SV model. However, there are some drawbacks associated with the stable diffusion models. In the estimation, we do not freely estimate the shape parameters of the stable Lèvy process. We instead specify several reasonable combinations of the shape and skewness parameters along a

rough grid.[21] Estimating the shape parameters freely may make an even better use of the stable process. Still, many relevant issues associated with using the stable distribution need to be explored in a systematic manner. For example, the non-existence of moments of second or higher order is a potential problem from an empirical point of view. Also, when using simulation-based estimation techniques, the value of the shape variable is found to be closely related to the size of the simulation, which introduces difficulties in model comparisons.

C. RS Models

The first two columns of Table 1.7c reports EMM estimates for the RS models that do not incorporate a level effect: the RS-in-σ model, and the RS-in-$\sigma + SV$ model. For these RS models, the p-values on the EMM objective function are higher than 5% but lower than 10%, providing mild evidence in support of the models. The fitted models imply strong mean reversion in the short rate. They indicate that the short rates are reverting to a positive long-run mean of 9.08% and 6.75%, respectively, which are substantially higher than the long-run mean estimates implied by the single-regime models.[22] The estimates of the regime dependent volatility parameters reveal that regime 1 is a high-volatility regime and regime 2 is a low-volatility regime. The two estimated regime switching probabilities, P_1 and P_2, exceed 0.90 and are similar to estimates reported by other authors. Notice that while the transition probability of

[21] This strategy has been used in Gallant et al. (1997) for their discrete-time SV model with Student-t errors. Further research on how the SNP model will encompass the Stable distributed errors is of strong interest.

[22] The substantially large difference of the long-run mean estimates in the single-regime models and RS models may suggest a regime-switching mean scenario, which could be left for future research.

staying in the low-volatility regime, P_1, are estimated similarly to those in previous empirical works, the estimates of P_2 (0.91 and 0.94) are slightly lower than what has been shown, implying less persistence of staying in the high-volatility regime for the dynamics of the short rate.

Without implementing the level effect, the RS-in-σ model only allows the conditional volatility to switch across two regimes; that is, any conditional heteroskedasticity can only be driven by switches of conditional volatility between two regimes. For such a simple model, the RS-in-σ model does a good job of modeling the volatility dynamics of short rates. It fits much better than the one factor Gaussian CKLS model, and slightly better than the one-factor non-Gaussian Stable CKLS model. The flexibility of incorporating two different levels of volatility is the main reason for the success of this simple RS model relative to many single-regime models. As argued by Gray (1996), the single-regime models treat volatility as being constant at an average level, in which case volatility estimates are uniformly too high during periods of low volatility and uniformly too low during periods of high volatility. Hence, the models fail to describe well the data in either regime.

Contrary to our expectations, the RS-in-$\sigma + SV$ model does not explain the dynamic behavior of short rates appreciably better than the RS-in-σ model even though it allows for an additional source of conditional heteroskedasticity driven by the volatility persistence beyond the regime switched conditional volatility. This is in contrast what Gray (1996) found with his RS-GARCH model. Using a likelihood ratio (LR) test to compare his RS-in-σ model with his RS-in-$\sigma + GARCH$ model, Gray (1996) showed that both the RS effect and the GARCH effect are important. Our EMM

estimation results imply that it is not necessary to incorporate the more complex RS-in-$\sigma + SV$ model.

The third column of Table 1.7c reports EMM estimates of the RS-in-$\sigma + Level$ model. This model, with an EMM objective function p-value of 0.29, fits much better than the RS models that do not incorporate the level effect. The RS-in-$\sigma + Level$ model can be described as a generalized CKLS model in which the conditional volatility switches between two very persistent regimes. Incorporating both a level effect and a RS effect, the RS-in-$\sigma + Level$ model provides the best performance in terms of fitting the volatility of short rates; it fits even better than the three-factor SV model, in which both the level effect and SV effect are implemented in the underlying structural model. It appears that the flexibility of having two volatility regimes and having the level effect picking up the remaining information is the main reason for the relative success of the RS-in-$\sigma + Level$ model over the single-regime models and the previous RS-in-σ models.

Compared with the ML estimates in Smith (2002), our EMM results are quite different in several respects. The estimated process is reverting to a lower long-term mean with a fast speed and smaller regime-dependent variances. Although the transition probability P_1 is similar, the estimate of P_2 at 0.89 is much lower than what has been shown by Smith (2002). In terms of the estimate for the level effect, the estimated conditional volatility in the RS-in-$\sigma + Level$ model is sensitive to the level of the short rates; that is, the level effect parameter is significantly different from zero. However, the magnitude of the estimated level effect, much lower than that reported in Smith (2002) at 0.92, is very similar as these in our multi-factor SV models. It appears that the

combination of the level effect with either a RS factor or a SV factor does not influence the importance of the level effect.

The last column of Table 1.7c shows results for the RS-in-$\sigma + Level + SV$ model. Characterized by combining all three effects of the level, RS, and SV effects within one model, the RS-in-$\sigma + Level + SV$ model is rejected by the EMM objective function at the significant level 5%. The score diagnostic t-ratios in Table 1.8b show that the score elements associated with the first, second, fourth Hermite polynomial elements and ARCH and GARCH coefficients are larger than two, which suggests that the RS-in-$\sigma + Level + SV$ model has trouble capturing the associated features as summarized by the 11117000 auxiliary model. A noticeable result for this most complex model specification is that the level effect has been almost squeezed out by the SV effect and the RS effect; its estimate is not significant from zero. Comparing the EMM results for the RS-in-$\sigma + Level + SV$ model with the other models, provides a way for addressing an important issue; that is, whether or not we need to include both RS and SV in the process of fitting the dynamics of the short rates. The answers from Gray (1996), Smith (2002), and many others are somewhat ambiguous due to the fact that traditional hypothesis testing procedures for evaluating the existence of Markov switching are nonstandard. For example, using the LR test, the statistical significance of the second regime cannot be tested using chi-square critical values because the parameters associated with the second regime are not identified under the null of a single regime. Although some extended tests have been developed for solving such kinds of difficulties, EMM provides a rather easy procedure to answer the issue by simply comparing the corresponding p-values for different non-nested model specifications.

From our estimation, it indicates that either a RS with level effect or an SV with level effect, but not both, are needed to adequately fit the data series of the short rate.

1.6. Conclusion

In this paper we develop a framework for evaluating and comparing the empirical fit of a number of discrete-time and continuous-time models for the US short rate. The models we consider include Gaussian diffusion models, non-Gaussian diffusion models with stable process, and different types of Markov RS models. A comprehensive model comparison is provided by utilizing the EMM estimation, which allows for ranking the non-nested model specifications. For the continuous-time models, we confirm the results from the existing empirical literature that the one-factor Gaussian diffusion model constitutes a poor candidate model for the short rate process. We find that a one-factor stable diffusion model shows stronger explanatory power to that of the one-factor Gaussian model, and that the multi-factor SV models (a two-factor SV model and a three-factor SV model) shows much better fitting performances. For the discrete-time RS models, we find that the simple RS-in-σ model, which allows the conditional variance to switch between regimes, describes the data surprisingly well. We also find that there are no fitting improvements of the extended RS-in-$\sigma + SV$ model over the RS-in-σ model, and of the extended RS-in-$\sigma + Level + SV$ model over the RS-in-$\sigma + Level$ model. These results suggest that either an SV effect or a RS effect, but not both, are needed for describing the data accurately. This point is consistent with the argument of Smith (2002), although his conclusion is much more informal and ambiguous. In summary, our multi-factor SV models and the RS-in-$\sigma + Level$ model

provide the overall best fits for the short rate process. The success of our two-factor SV model is opposite to the general belief exiting in the literature that two factors are not enough to accommodate the complex process of short rates. Figure 1.7 displays representative simulated paths from three preferred models. Relative to the actual interest rate series in Figure 1.2, the three simulation series are capable of generating some extreme volatile periods as the monetary experiment experience, and share qualitative features with the actual interest rate data.

We also provide insights on the measurement of one of the important features of the US short rates, the level effect. Our finding shows that the level effect is similarly estimated a bit higher than 1/2 in the preferred multi-factor SV models and the RS-in-$\sigma + Level$ model, which is consistent with the finding in AL (1997). Although the corresponding estimate obtained from the RS-in-$\sigma + Level + SV$ model is significantly weakened, the estimated parameter is not significantly different from zero. Our estimations imply that the estimated level effect is relatively robust to the sample used for estimation; it may be spuriously low or high for misspecified models that fail to capture the time-varying and heteroskedastic behavior of the short rates.

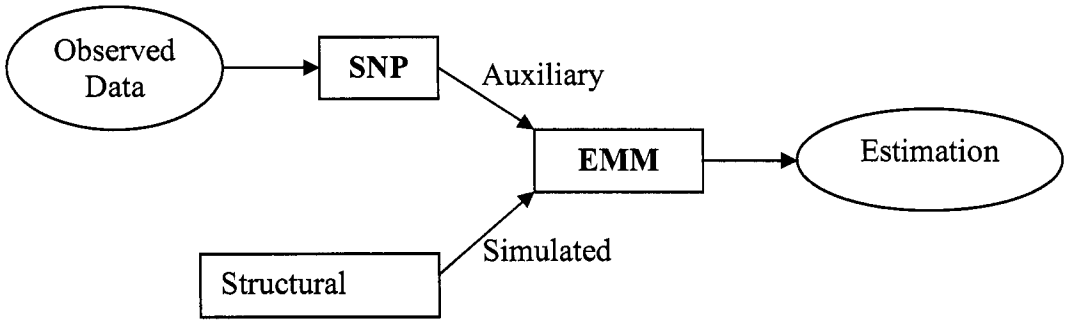
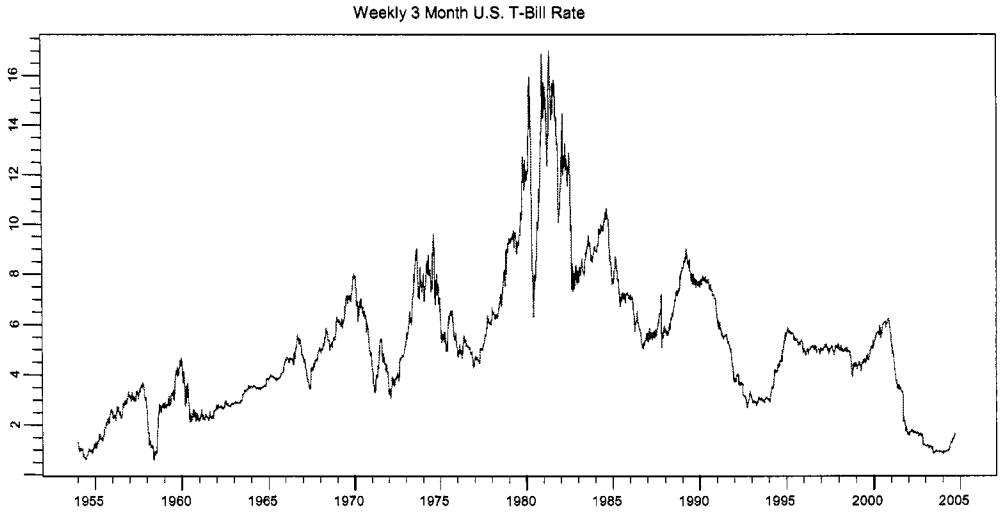


Figure 1.1: Procedures of EMM Methodology

EMM procedure consists of two steps (1) the projection step, which is accomplished by projecting the data onto the SNP model, and (2) the estimation step, in which structural parameters are extracted from the summary of the data by minimizing the chi-squared criterion.

(A)



(B)

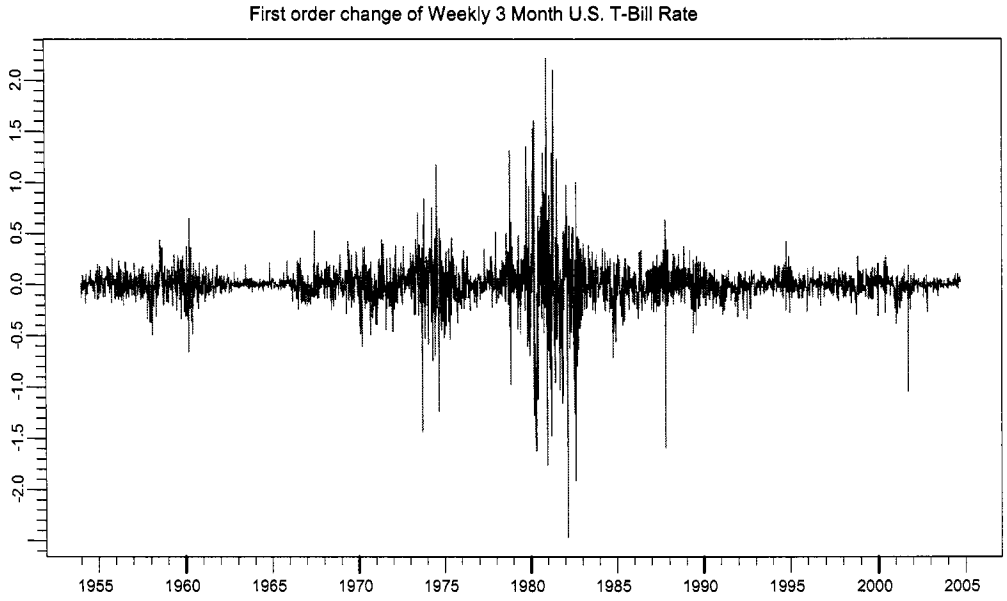
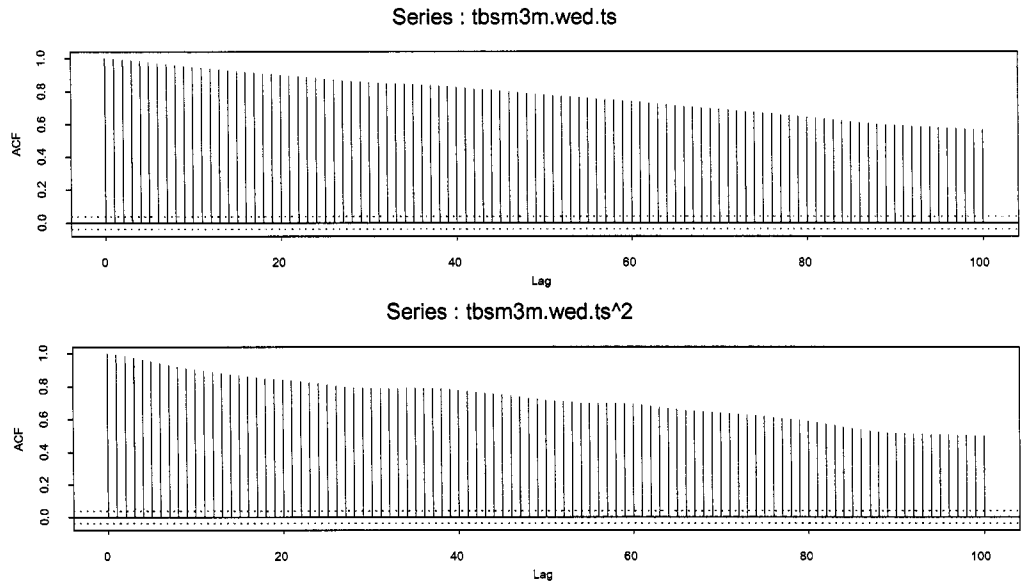


Figure 1.2: Time Series Plots

There are 2648 weekly observations of the 3-month T-Bill rates, ranging from January 4, 1954 to September 24, 2004. The raw data (percent) is plotted in panel (A); the first order difference of the raw data is presented in panel (B).

(A)



(B)

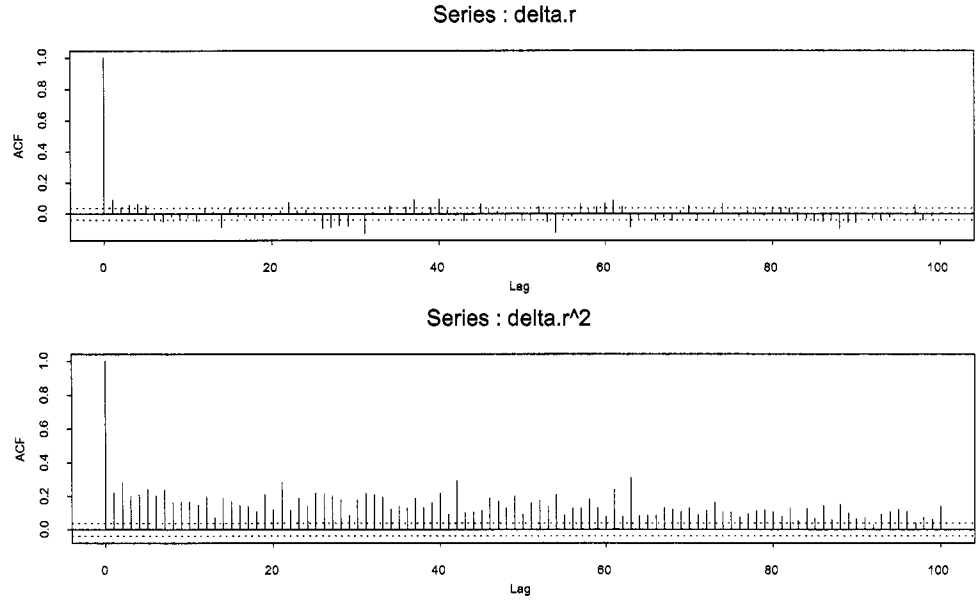
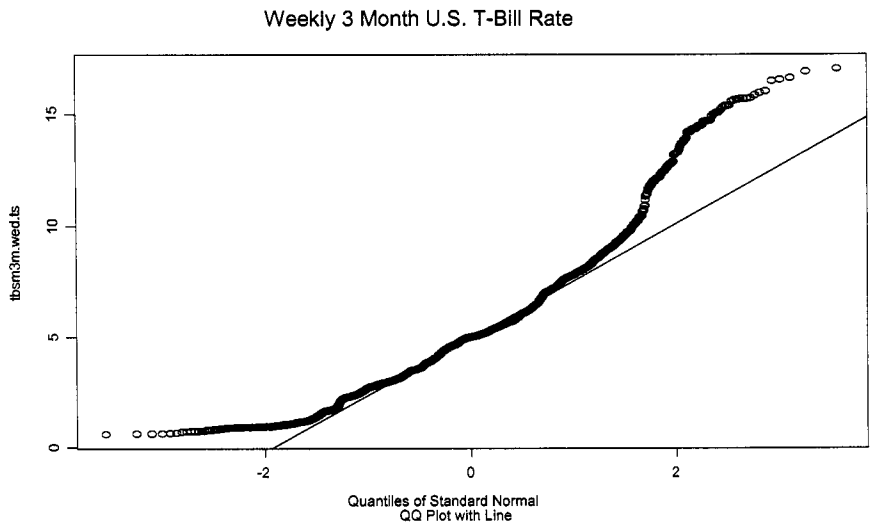


Figure 1.3: Autocorrelation Plots

The ACF plots for the raw data (percent) and the squared series are given in panel (A); the ACF plots of the first order difference of the raw data and its squared series are presented in panel (B).

(A)



(B)

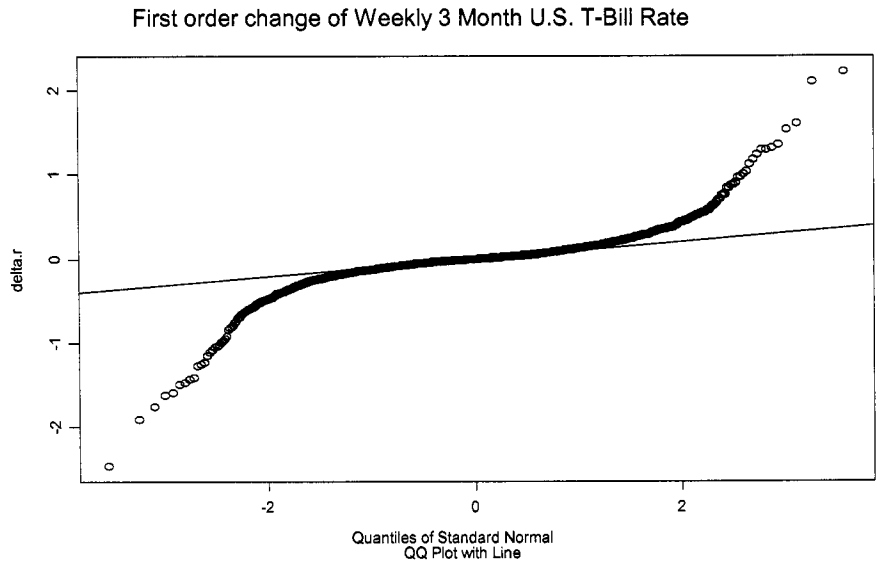


Figure 1.4: QQ Plots with 45° QQ Line

The qq-plot for the raw data (percent) is given in panel (A); the qq-plot of the first order difference of the raw data is presented in panel (B).

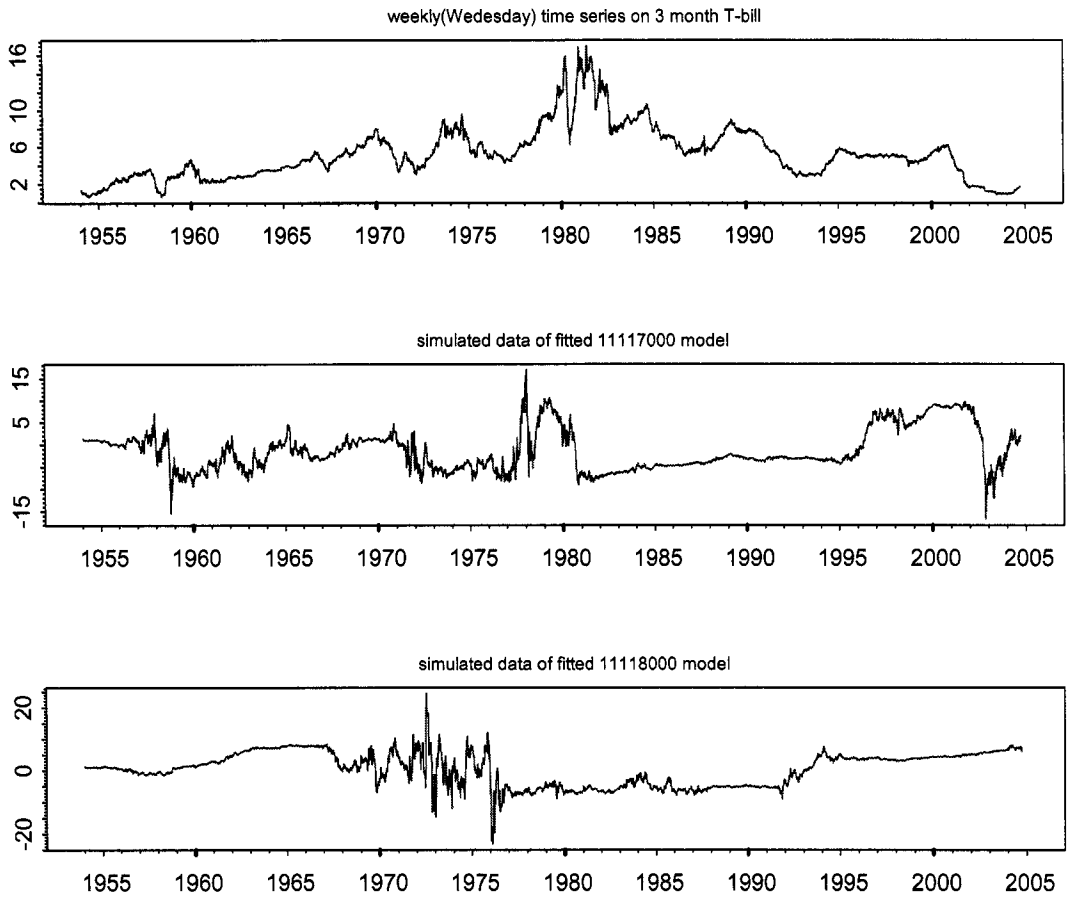
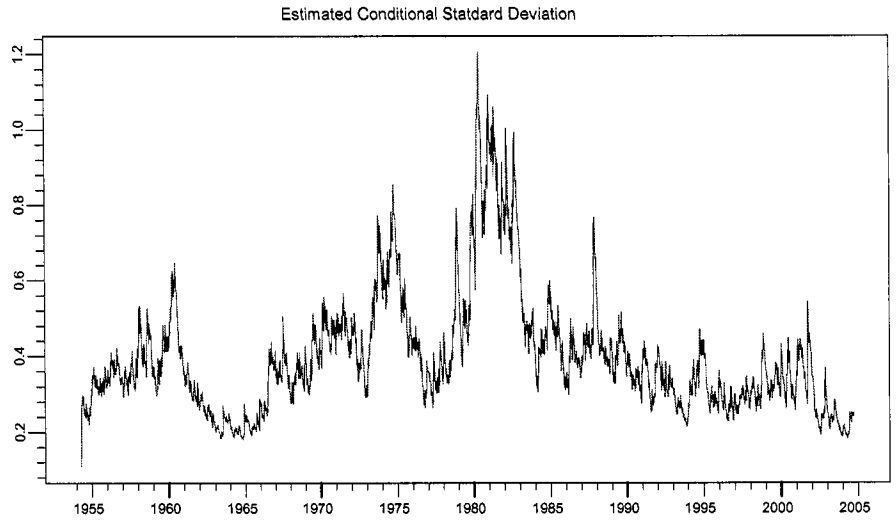


Figure 1.5: Simulated data from Fitted SNP Models

(A)



(B)

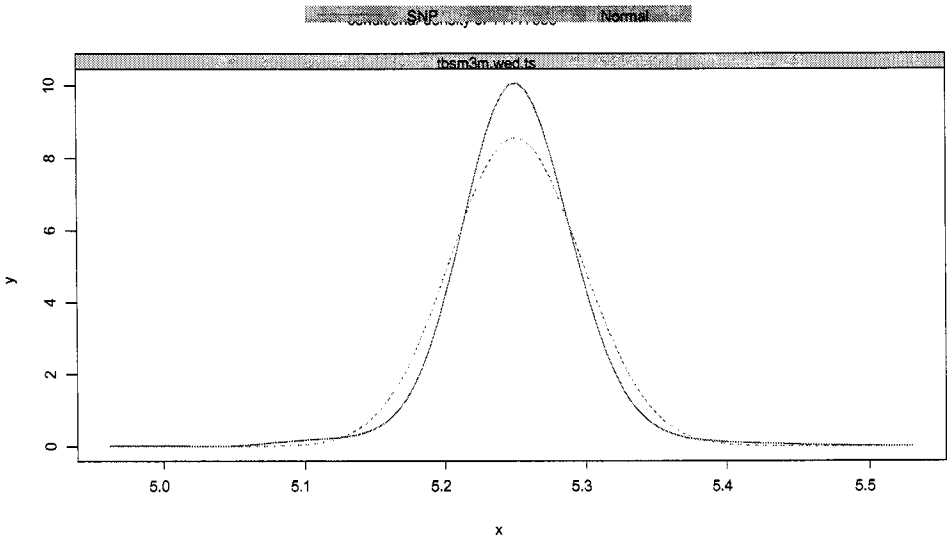
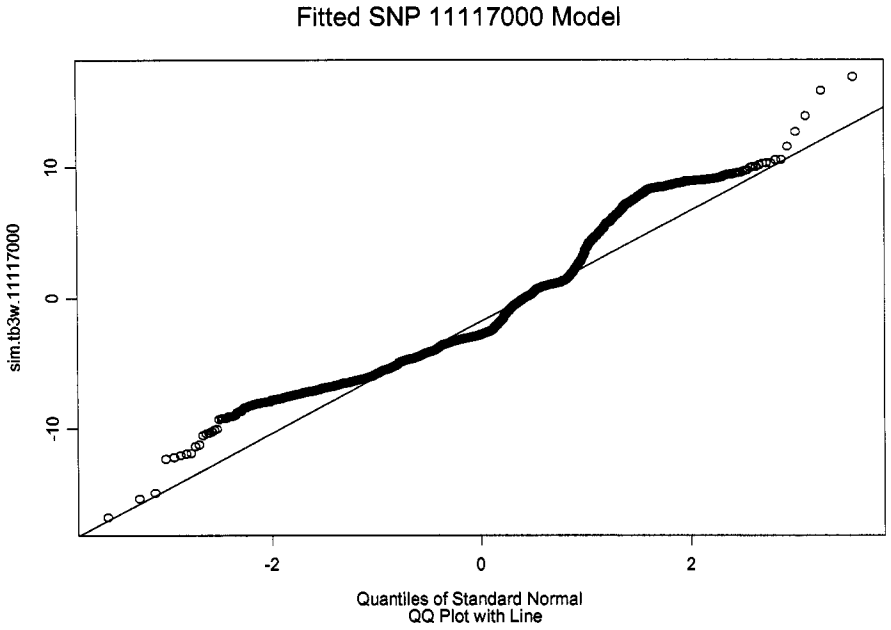


Figure 1.6a: Diagnostic Tests for SNP Model 11117000 I

The panel (A) gives the estimated conditional volatilities of the data, which is persistent and volatile; the panel (B) shows the conditional density, which is peaked in the center with heavy tails.

(C)



(D)

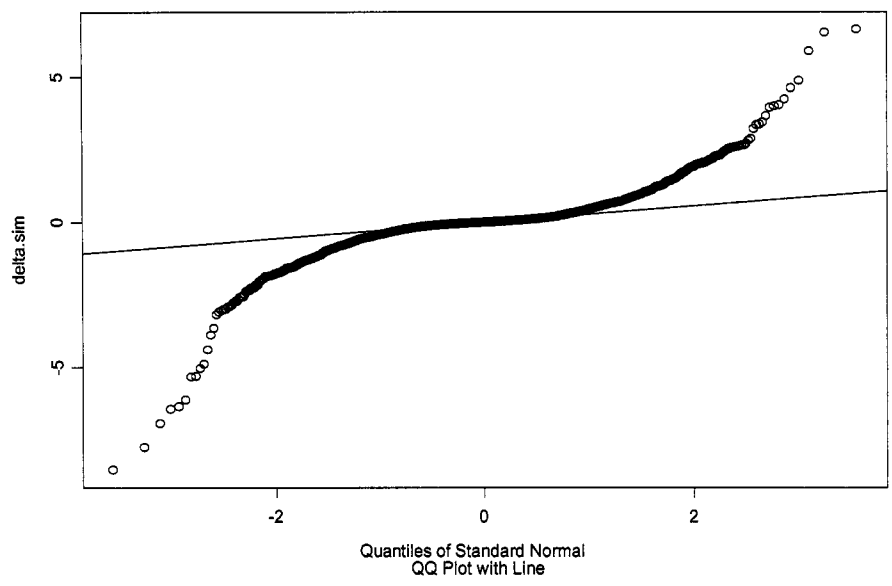
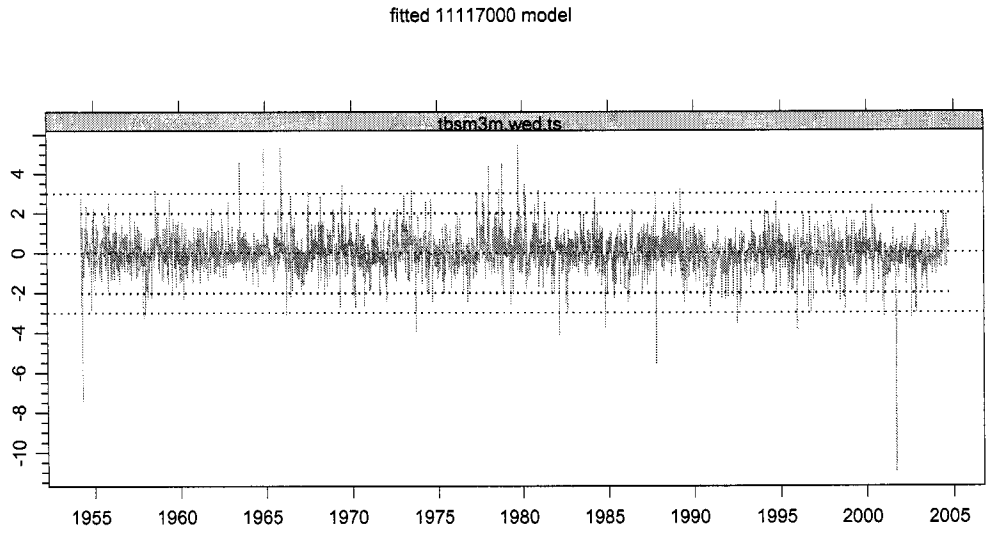


Figure 1.6b: Diagnostic Tests for SNP Model 11117000 II

The panel (C) gives the qq-plot of the simulated series from the preferred SNP model 1117000; the panel (D) shows the qq-plot of the change of the simulated series.

(E)



(F)

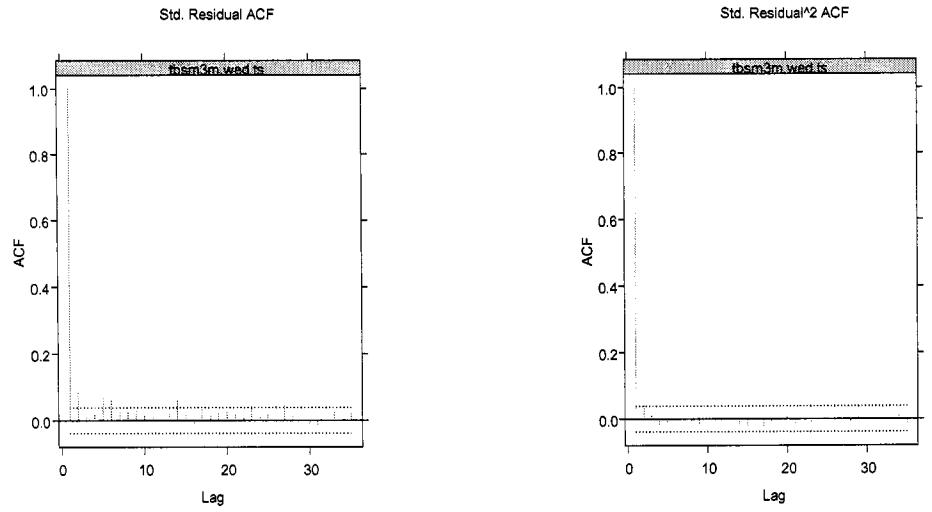
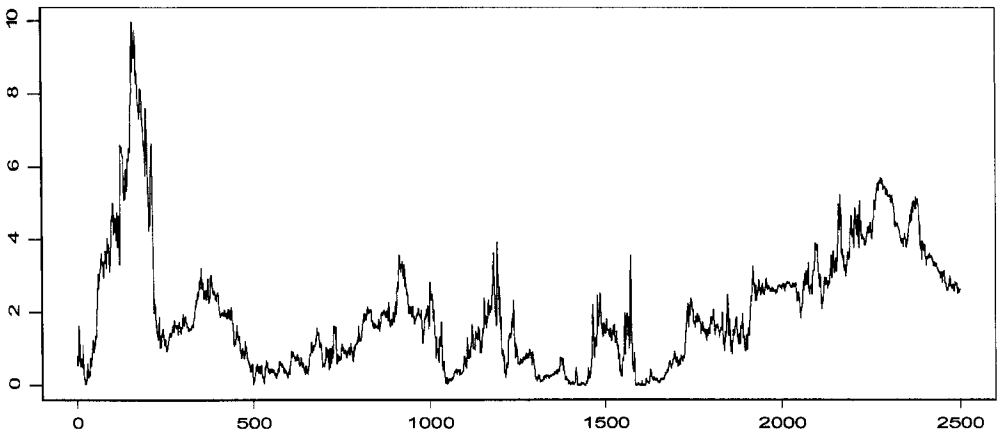


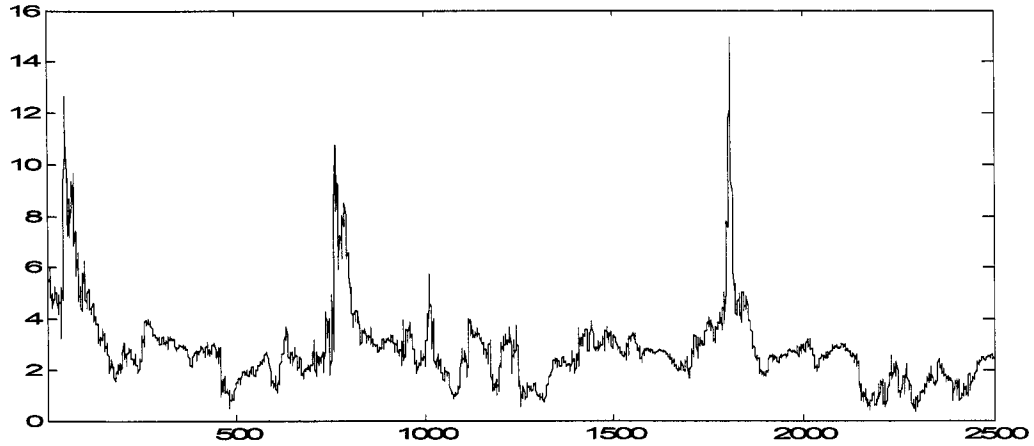
Figure 1.6c: Diagnostic Tests for SNP Model 11117000 III

The panel (E) represents the standardized residuals, which seems to resemble a Gaussian white noise process. Lastly, the panel (F) provides the autocorrelation plot for the residuals and the squared residuals, implying no significant autocorrelation for both of them.

(A) Simulation from the fitted two-factor SV model



(B) Simulation from the fitted three-factor SV model



(C) Simulation from the fitted RS-in-σ + *Level* model

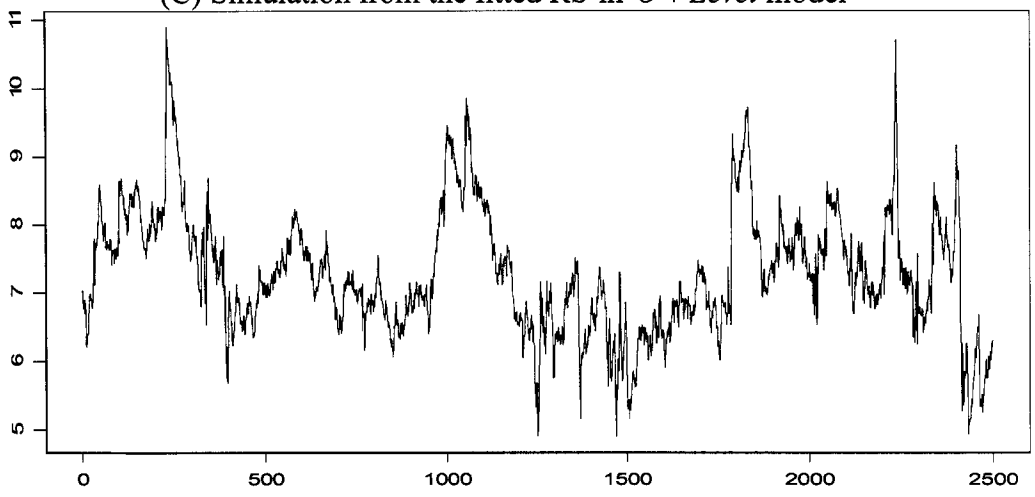


Figure 1.7: Plots of Simulations from Preferred Models

Table 1.1: SNP Tuning Parameters

A useful taxonomy of SNP models is defined by putting certain restrictions on the tuning parameters, according to Gallant and Tauchen (1997).

Parameter Setting	Characterization of
$L_u = 0, L_g = 0, L_r = 0, L_p \geq 0, K_z = 0, K_x = 0$	iid Gaussian
$L_u > 0, L_g = 0, L_r = 0, L_p \geq 0, K_z = 0, K_x = 0$	Gaussian VAR
$L_u > 0, L_g = 0, L_r = 0, L_p \geq 0, K_z > 0, K_x = 0$	Semiparametric VAR
$L_u \geq 0, L_g = 0, L_r > 0, L_p \geq 0, K_z = 0, K_x = 0$	Gaussian ARCH
$L_u \geq 0, L_g = 0, L_r > 0, L_p \geq 0, K_z > 0, K_x = 0$	Semiparametric ARCH
$L_u \geq 0, L_g > 0, L_r > 0, L_p \geq 0, K_z = 0, K_x = 0$	Gaussian GARCH
$L_u \geq 0, L_g > 0, L_r > 0, L_p \geq 0, K_z > 0, K_x = 0$	Semiparametric GARCH
$L_u \geq 0, L_g \geq 0, L_r \geq 0, L_p > 0, K_z > 0, K_x > 0$	Nonlinear nonparametric

Table 1.2: Descriptive Statistics

The statistics summary is given in Panel (A) for the raw data, which are 2648 weekly observations of the 3-month T-Bill rates, ranging from January 4, 1954 to September 24, 2004. The panel (B) shows the statistics summary for the change of the raw data.

(A)

Sample Quantiles	Min: 0.6	1Q: 3.18	Median: 4.99	3Q: 6.67	Max:17.01
Sample Moments	Mean: 5.246	Std. Dev. : 2.849	Skewness: 1.065	Kurtosis: 4. 712	

(B)

Sample Quantiles	Min: -2.47	1Q: -0.07	Median: 0	3Q: 0.077	Max:2.22
Sample Moments	Mean:1.5e-4	Std. Dev. : 0.236	Skewness: -0.522	Kurtosis: 24.81	

Table 1.3: SNP Fitting Strategy

The SNP score generator has a leading (G)ARCH term with L_u lags in conditional mean. The standardized innovation has a normal density stretched by a squared Hermite polynomial with degree of K_z. Similarly, the coefficient of the z-polynomial may depend on the lagged observations through a K_x degree polynomial. For univariae SNP density, the interaction polynomial terms, I_z and I_x, are ignored. The fitting strategy is shown by the following diagram.

VAR-ARCH leading:

```
   10010000 →→→→ 10110000
→ 20010000       10210000
   30010000       …
   …              10810000→ 10814000
   70010000                 10815000
   80010000                 …
                            10818000 → 10818010; 10818020
```

VAR-GARCH leading:

```
10010000 → 11110000
           11114000
           …
           11117000  → 11117010; 11117020 (check conditional heterogeneity)
           11118000  → 11118010; 11118020
```

Table 1.4: SNP Estimation and Selection

The continued table reports the choice of SNP density and the BIC value, based on which we choose our preferred specification. We find that a VAR-GARCH model 11117000 is the BIC preferred model for T-Bill rates using the searching strategy specified in Table 3.

ARCH-Leading	L_u	L_g	L_r	L_p	K_z	I_z	K_x	I_x	BIC
10110000	1	0	1	1	0	0	0	0	-1.377
10210000	1	0	2	1	0	0	0	0	-1.4851
10310000	1	0	3	1	0	0	0	0	-1.5307
10410000	1	0	4	1	0	0	0	0	-1.5618
10510000	1	0	5	1	0	0	0	0	-1.5729
10610000	1	0	6	1	0	0	0	0	-1.5836
10710000	1	0	7	1	0	0	0	0	-1.5866
10810000	**1**	**0**	**8**	**1**	**0**	**0**	**0**	**0**	**-1.587**
10814000	1	0	8	1	4	0	0	0	-1.6238
10815000	1	0	8	1	5	0	0	0	-1.6221
10816000	**1**	**0**	**8**	**1**	**6**	**0**	**0**	**0**	**-1.6301**
10817000	1	0	8	1	7	0	0	0	-1.6287
10818000	1	0	8	1	8	0	0	0	-1.6285
10816010	1	0	8	1	6	0	1	0	-1.6247
10816020	1	0	8	1	6	0	2	0	-1.6152
GARCH-Leading	L_u	L_g	L_r	L_p	K_z	I_z	K_x	I_x	BIC
11110000	1	1	1	1	0	0	0	0	-1.5957
11114000	1	1	1	1	4	0	0	0	-1.6374
11115000	1	1	1	1	5	0	0	0	-1.6359
11116000	1	1	1	1	6	0	0	0	-1.6344
11117000	**1**	**1**	**1**	**1**	**7**	**0**	**0**	**0**	**-1.6407**
11118000	1	1	1	1	8	0	0	0	-1.6395
11117010	1	1	1	1	7	0	1	0	-1.6325
11117020	1	1	1	1	7	0	2	0	-1.622
11118010	1	1	1	1	8	0	1	0	-1.6298
11118020	1	1	1	1	8	0	2	0	-1.618

Table 1.5: Parameter Estimates of Projected SNP Density

This preferred SNP model of 11117000 is a GARCH (1,1) with a nonparametric error density represented as a seven-degree Hermite expansion where the Hermite coefficients are state independent. The auxiliary model and the conditional density 1117000 are given by

$$y_t = \mu_0 + b_1 x_{t-1} + \sigma_t z_t$$

$$\sigma_t = \rho_0 + p_1 \mid y_{t-1} - \mu_{x_{t-2}} \mid + g_1 \sigma_{t-1}$$

$$f(y_t \mid x_{t-1}, \theta) = \left[P(z_t, x_{t-1})\right]^2 N(y_t; \mu_0 + b_1 y_{t-1}, \sigma_t^2)$$

$$\text{with } P(z_t, x_{t-1}) = \sum_{i=0}^{K_z=7} a_i z_t^i$$

Parameter	Estimate	Standard Error	T-statistics
Hermite a_0	0.0000	0. 0000	0.000
Hermite a_1	-0.0431	0. 0330	-1.3072
Hermite a_2	-0.3073	0. 0215	-14.3144
Hermite a_3	0.0299	0. 0189	1.5790
Hermite a_4	0.0523	0. 0059	8.8977
Hermite a_5	-0.0071	0. 0033	-21358
Hermite a_6	-0.0021	0. 0004	-5.3421
Hermite a_7	0.0004	0. 0002	2.8532
Mean μ_0	0.0021	0. 0012	1.7891
Mean b_1	0.9996	0. 0009	1106.6166
ARCH ρ_0	0.0017	0. 0001	12.3629
GARCH p_1	0.2463	0. 0150	16.4166
GARCH g_1	0.8517	0. 0072	117.8965
BIC: -1.6418	HQ: -1.6503	AIC: -1.6551	Log L: 4371.665

Table 1.6: SNP Models used by Selected Papers for Financial Data
This table lists a number of SNP models that have been utilized for EMM estimations in selected papers with applications of the stochastic volatility modeling, the term structural dynamics and long memory study for the interest rates and equity returns.

Selected Papers	SNP model	Application
Andersen and Lund (1997)	51116000*	Stochastic volatility models of the short-term interest rate.
Bansal and Zhou (2000)	10514300	Term structure models using the bivariate dynamics of the yields on the six-month bill and the five-year note.
Ahn et al. (2003)	11114300	Term structure models using the bivariate dynamics of the yields on the six-month bill and the three-year note.
Dai and Singleton (2000)	10214000	Affine term structure models using the swap rates of maturities from six months to ten years.
Andersen et al. (2002)	01118000*	Stochastic volatility models of the S&P 500 Index return.
Chernov et al. (2000)	11118000	Stochastic volatility models and jump diffusion models of the Dow Jones Industrial Average Index return.
Liu (2000)	0025018000**	Long memory of equity returns.

* Andersen and Lund (1997) used an SNP model with EGARCH (1,1), instead of Level-GARCH (1,1), as the leading term.
** This SNP model is a VAR(0) with ARCH(25) conditional variance and the nonparametric error is represented by a stage-independent Hermite Polynomial of degree 18.

Table 1.7a: EMM Model Estimations I

The EMM estimations are given for the one-factor Gaussian diffusion model based on the CKLS model (CKLS-N) and corresponding two-factor and three-factor SV models (SV2 and SV3), which are laid out in section 3.1. The CKLS-N model refers to the model (1.9); the SV2 model refers to the one (1.11); and the SV3 model refers to the one (1.12) with three stochastic factors.

Parameter	CKLS*	CKLS-N	SV2	SV3
ϕ_0	0.0408 (0.022)	0.4818 (0.02593)	0.8428 (0.2669)	-
$-\phi_1$	-0.5921 (0.382)	-0.1927 (0.01830)	-0.2956 (0.1816)	-1.0947 (0.8463)
ϕ_0 / ϕ_1 **	0.0690 (-)	2.5003 (-)	2.8512 (-)	- (-)
γ	1.4999 (0.252)	0.3076 (0.05062)	0.6659 (0.1163)	0.5167 (0.0755)
σ	1.6704 (2.169)	1.3624 (0.19325)	-	-
ω_0	-	-	-0.5912 (0.5122)	-1.4491 (0.3003)
ω_1	-	-	-0.5629 (0.1983)	-0.9250 (0.1990)
ξ	-	-	1.7765 (0.0474)	2.5433 (0.0959)
υ_0	-	-	-	2.5494 (0.7520)
υ_1	-	-	-	-0.9801 (16.5030)
ζ	-	-	-	0.4715 (2.1140)
χ^2	-	50.47	10.35	5.55
$p-value$	-	3.32e-08	0.1107	0.2357
$d.o.f$ ***	-	8	6	4

* The model of Chan et al (1992) with monthly short-term interest rates over period of 6/1964 to 12/1989.
** The fitted long-run reverting mean.
*** The "d.o.f" stands for the "degree of freedom".

Table 1.7b: EMM Model Estimations II

The EMM estimations are given for the non-Gaussian diffusion models based on the CKLS model, which is specified by

$$dr_t = (\phi_0 - \phi_1 r_t)dt + \sigma r_t^{\gamma} dL_t = k_r(\mu_r - r_t)dt + \sigma r_t^{\gamma} dL_t$$

where the L_t is stable Lèvy process with shape variable α, and skewness variable β. The CKLS-N model where L_t is Wiener process is listed for the comparison with the CKLS-S(α, β) models.

Parameter	CKLS*	CKLS-N	CKLS-S (1.8,0)	CKLS-S (1.9,0)	CKLS-S (1.95,0)	CKLS-S (1.9,0.1)
ϕ_0	0.0408 (0.022)	0.4818 (0.02593)	0.9043 (0.05396)	0.7546 (0.1272)	0.7202 (0.06407)	1.2154 (0.08063)
$-\phi_1$	-0.5921 (0.382)	-0.1927 (0.01830)	-1.2370 (1.237)	-0.5325 (0.373)	-0.4502 (0.12936)	-4.7425 (0.31014)
ϕ_0 / ϕ_1	0.0690 (-)	2.5003 (-)	0.7310 (-)	1.4171 (-)	1.5997 (-)	0.2563 (-)
γ	1.4999 (0.252)	0.3076 (0.05062)	0.3727 (0.01508)	0.3664 (0.0177)	1.0675 (0.06533)	0.4905 (0.03837)
σ	1.6704 (2.169)	1.3624 (0.19325)	-0.5207 (0.19325)	0.9694 (0.0171)	1.0675 (0.02534)	0.8965 (0.05622)
χ^2	-	50.47	82.19	33.39	23.89	15.8
$p-value$	-	3.32e-08	1.765e-14	5.238e-05	1.195e-04	0.0454**
$d.o.f$	-	8	8	8	8	8

* The model of Chan et al (1992) with monthly short-term interest rates over period of 6/1964 to 12/1989.
** The corresponding p-value with degree of freedom of six is 0.0149.

Table 1.7c: EMM Model Estimation III

The following estimations are given for the different types of Markov Regime Switching (RS) models, which are laid out in section 3.3. The first two models are OU-based RS models, given by equations of (1.14) and (1.16) respectively. The last two models are CKLS-based, given by equations of (1.17) and (1.18) respectively.

Parameter	OU Based		CKLS Based	
	RS-in-σ	RS-in-σ + SV	RS-in-σ + Level	RS-in-σ + Level + SV
ϕ_0	0.2580 (0.0336)	0.2240 (0.6623)	0.1769 (0.5818)	0.4027 (0.3227)
$-\phi_1$	-0.0284 (0.1640)	-0.0332 (0.5006)	-0.0285 (0.2976)	-0.0408 (1.0771)
ϕ_0 / ϕ_1	9.0845 (-)	6.7470 (-)	6.2070 (-)	9.8701 (-)
γ	-	-	0.5076 (0.1266)	0.0063 (4.5077)
σ_1	0.1472 (0.0628)	-	0.0389 (0.0914)	-
σ_2	0.4613 (0.0556)	-	0.1400 (0.0874)	-
ω_{01}	-	-2.1187 (1.6037)	-	-2.8481 (0.3137)
ω_{02}	-	-1.0078 (0.6564)	-	-1.3966 (0.1915)
ω_1	-	-0.5326 (1.5275)	-	-0.1570 (0.4406)
ξ	-	0.2902 (0.3272)	-	0.2210 (0.2212)
P_1	0.98	0.98	0.98	0.98
P_2	0.91	0.94	0.89	0.94
χ^2	11.01	8.94	6.53	5.55
$p-value$ *	0.0881	0.0626	0.2916	0.0253
$d.o.f$	6	4	5	3

* The p-values are calculated based on the degree of freedom equal to the $d.o.f$ less two.

Table 1.8a: Models Diagnostic T-Ratios I

The adjusted t-rations* are reported for different model specifications based on the same score generator (11117000), for which the parameters refers to the following equations. The adjusted t-ratios are testing whether the fitted sample moments are equal to zero, as predicted by population moments of the SNP density.

$$y_t = \mu_0 + b_1 x_{t-1} + \sigma_t z_t, \text{ where } \sigma_t = \rho_0 + p_1 \mid y_{t-1} - \mu_{x_{t-2}} \mid + g_1 \sigma_{t-1},$$

$$f(y_t \mid x_{t-1}, \theta) = \left[P(z_t, x_{t-1})\right]^2 N(y_t; \mu_0 + b_1 y_{t-1}, \sigma_t^2) \quad \text{with} \quad P(z_t, x_{t-1}) = \sum_{i=0}^{K_z=7} a_i z_t^i$$

Parameter	CKLS-N	CKLS-S(1.95, 0)	CKLS-S(1.9, 0.1)	SV2	SV3
Hermite a_1	-1.5845	-1.2618	0.03594	-0.3861	0.5721
Hermite a_2	**-2.4841**	-0.7672	-0.5429	0.08518	-0.8662
Hermite a_3	**-2.7178**	**-3.4528**	-1.884	-1.5801	-0.8818
Hermite a_4	**2.0221**	0.5788	-0.3106	1.05861	-0.0750
Hermite a_5	**-2.0172**	**-3.0735**	-1.9019	-1.2937	-0.7872
Hermite a_6	**2.7662**	1.1112	0.09164	1.30066	0.2805
Hermite a_7	-0.5218	-1.4579	-1.324	-0.2149	-0.1253
Mean μ_0	1.1283	0.8312	0.27349	1.71367	1.7819
Mean b_1	**-2.4661**	-1.6667	0.13483	-1.1658	-0.5640
GARCH ρ_0	-1.6397	0.3467	-1.8993	0.6306	0.5843
GARCH p_1	**-2.1221**	0.4447	-1.6631	0.2647	0.3493
GARCH g_1	-1.9152	0.2685	-1.6	-0.3861	0.4326
$p-value$	3.32e-08	1.20e-04	0.0454	**0.1107**	**0.2357**

* According to Gallant and Tauchen (2000), the unadjusted t-ratios are biased downward.

Table 1.8b: Models Diagnostic T-Ratios II

The t-rations are reported for different model specifications based on the same score generator (11117000), for which the parameters refers to the following equations. The t-ratios are testing whether the fitted sample moments are equal to zero, as predicted by population moments of the SNP density.

$$y_t = \mu_0 + b_1 x_{t-1} + \sigma_t z_t, \text{ where } \sigma_t = \rho_0 + p_1 \mid y_{t-1} - \mu_{x_{t-2}} \mid + g_1 \sigma_{t-1},$$

$$f(y_t \mid x_{t-1}, \theta) = \left[P(z_t, x_{t-1})\right]^2 N(y_t; \mu_0 + b_1 y_{t-1}, \sigma_t^2) \quad \text{with} \quad P(z_t, x_{t-1}) = \sum_{i=0}^{K_z=7} a_i z_t^i$$

Parameter	OU Based		CKLS Based	
	RS-in-σ	RS-in-σ + SV	RS-in-σ + Level	RS-in-σ + Level + SV
Hermite a_1	-0.7338	0.6262	0.7327	**-2.1586**
Hermite a_2	-1.7376	**-2.4572**	-1.5314	**-5.6508**
Hermite a_3	-0.7855	-0.2299	0.3636	-0.4873
Hermite a_4	-1.0537	-1.6070	-1.1830	**-2.5206**
Hermite a_5	-0.6400	-0.3290	-0.0961	-0.0609
Hermite a_6	-0.8775	-0.2241	-0.9017	-0.7921
Hermite a_7	-0.8047	0.3008	-0.9423	0.1670
Mean μ_0	-0.3481	1.7798	0.2709	-0.8511
Mean b_1	0.4950	-0.9013	0.9051	1.4379
GARCH ρ_0	-1.6486	-0.4648	-1.5102	-1.7896
GARCH p_1	-1.5090	-0.7646	-0.5848	**-2.7740**
GARCH g_1	-1.5210	-0.4723	-0.5863	**-2.5500**
$p-value$	0.0881	0.0626	0.2916	0.0253

Chapter II: A Filter Study of Volatility Models using EMM Estimation for U.S. Short Rate

2.1. Introduction

Due to the essential role of volatility modeling for asset pricing, academic researchers and investment institutions have devoted significant resources to developing and testing models for the volatility dynamics of financial asset prices. The econometric challenge for estimating volatility models is that traditional estimation methods are often infeasible due to the fact that volatility is not directly observable. After decades of effort, selection of the appropriate model and estimation methodology is still openly debated.[23] Among a wide range of methods[24] available for estimating partially observable systems, the efficient method of moments (EMM) (Gallant and Tauchen, 2001, 2002) has shown significant estimation flexibility and efficiency[25] and has been utilized in many financial applications.

While there is a sizable literature that uses EMM to estimate parameters of volatility models, little attention has been paid within this literature to volatility tracking and filtering after the estimation. The importance of recovering the latent volatility through filtering and smoothing is two-fold. First, forecasting future volatilities is vitally important for asset pricing, and volatility filtering is a crucial step for effective

---

[23] See Andersen et al. (1999).

[24] For example, generalized method of moments (GMM) (Melino and Turnbull, 1990), simulated method of moments (SMM) (Duffie and Singleton, 1989), quasi-maximum likelihood (QML) (Harvey et al., 1994), Bayesian Markov-Chain Monte Carlo analysis (MCMC) (Jacquier et al., 1994), Bayesian importance-sampling Monte Carlo (Geweke, 1994), a unified Markov-Chain Monte-Carlo sampling-based framework for Bayesian and maximum likelihood inference (Kim and Nelson, 1999), simulation-based maximum likelihood (SML) (Danielsson, 1994), maximum likelihood Monte Carlo (MCL) (Sandmann and Koopman, 1998), and direct maximum likelihood through recursive numerical integration (ML) (Fridman and Harris, 1998).

[25] The efficiency of EMM has been proven to converge to that of the maximum likelihood (ML) estimator under some regularity conditions.

forecasts. There are two types of volatility measures that have been used for volatility forecasting. A standard one is the *implied volatility* of a specific asset, which is typically obtained from the Black-Scholes option pricing model by applying a suitable pricing formula backwards. The implied volatility can then be used to price other securities with the same underlying asset. However, the assumptions adopted by the Black-Scholes model are obviously not consistent with actual financial markets. In addition, the volatilities implied from different securities for the same asset should be the same if the asset pricing models were correct. However, as empirical evidence has shown, the *implied volatility* for an asset can depend upon many factors, such as the strike or expiration of options, from which they are obtained. Another approach is to apply time series techniques for the historical series; volatility estimated in this way is called the *historical volatility*. The advantage of the historical volatility over the implied volatility is that the former concept reflects the actual market fluctuation of the volatility for a specific asset during a period of time; the implied volatility just represents a market consensus. However, even the ARCH/GARCH model, the most famous econometric model, has preformed poorly in forecasting volatilities. With the availability of high-frequency data and the continued advance of computational tools, there appear to be two ways to improve forecasting power using historical volatility. One way treats volatility as observed and uses intra-period squared returns to get the so-called *realized volatility*, which can be made arbitrarily close to the underlying integrated volatility by sampling returns at increasingly higher intra-daily periods.[26] While the introduction of realized volatility is promising in many forecasting contexts, high-frequency data may

[26] See Andersen et al. (2005) for details.

not be readily available for all financial assets; for example, interest rate data is only publicly-available on limited frequencies although there are some public data bases for intraday bond prices. Therefore, a different approach is to treat volatility as latent in the sense that it can be filtered after estimation. That is, parameters for a given model are estimated first, and then the latent volatility process is filtered and the sequential forecast for any desired sampling frequency could be obtained. This paper applies this approach to the US short rate using the EMM estimation results.

Secondly, examining the volatility filtering results for a given estimation will provide complementary information on model selections for EMM estimation. Without theoretical guidance about the appropriate model specification in a specific setting, the competition among various volatility models is genuine; each model tries to fit the data better than the other. Although EMM estimation has shown to be efficient for many types of models[27], it compares and ranks different model specifications simply based on the p-value of EMM test statistics that are asymptotically χ^2 distributed.[28] Our Monte Carlo study for a regime switching (RS) model in Chapter 3 showed that misspecified structural models may pass the EMM specification test. This fact implies that ranking models using p-values alone is not enough to compare models, especially non-nested models, and more informative tests may be needed to examine models, especially those that can not be rejected by EMM. For volatility modeling, filtering results offer a valuable test of evaluating the adequacy of a candidate model. That is, a successful

---

[27] Andersen et al. (1999) presented a Monte Carlo study of a stochastic volatility (SV) model using EMM; we showed that EMM estimation can also accommodate general extensions to the functional form and distributional assumption of regime switches in Chapter 3.

[28] The ML-type approach ranks the goodness-for-fit of different volatility models based on values of the optimal log-likelihood.

model among many should not only provide a good fit to the data (implying a higher p-value), but also present reasonable filtered volatilities that could resemble those of the observed data and business cycles of the underlying economy. This filtering examination may play a crucial role when EMM specification tests provide less information for distinguishing between models that have similar fitting performances or have similar p-values. For example, we estimated a variety of volatility models for fitting the US short rate in Chapter 1, from which a two-factor and a three-factor SV models and several RS models cannot be rejected by EMM at a 5% significant level. The filtering performance of the corresponding volatility could help provide more insight about the competition between the SV model and the RS model.

Therefore, the aim of this chapter is to comparatively evaluate the performance of a variety of volatility models in terms of the filtered and smoothed estimates of the latent volatility. For such a filtering task, Gallant and Tauchen (1998) proposed, after the projection step and estimation step of EMM, a reprojection step, which is a simulation-based nonlinear Kalman filtering technique. In this chapter, we instead apply appropriate algorithms from the literature on optimal filtering. Specifically, we first estimate the parameters of candidate volatility models for the US short rate using EMM. With the parameter estimates in hand, we then use optimal filtering algorithms to infer the unobserved volatilities from simulated data and then on a series of the US short rate. The first two models we consider are continuous-time stochastic volatility models where the latent volatility is continuous-valued. We apply particle filtering algorithms to these models, which were first introduced by Gordon, Salmon and Smith (1993) and extended in Pitt and Shephard (1999). Particle filters have been shown to accurately

estimate unobservable processes in nonlinear, non-Gaussian state space models. They have been considered for diffusion-based models by Johannes et. al. (2006) and more recently by Fearhead et. al. (2006). We demonstrate the reliability of different particle filtering algorithms by applying them on simulated data. Then we evaluate each volatility model in terms of their ability to track and filter volatility by applying them to the U.S. short interest rate. The second set of models we consider have latent states that are discrete-valued. In this case, the optimal filter is the discrete-state hidden Markov model (HMM) filter first developed by Baum and Petrie (1968).

The remainder of the chapter is organized as follows. A detailed description of the particle filtering methodology is provided in Section 2. Section 3 briefly summarizes the EMM methodology, the data, the models under consideration, and their EMM estimation results. Section 4 reports the empirical results using the simulated data and the US short rate and section 5 concludes.

2.2. Particle Filtering Algorithms[29]

A dynamic system consists of two models: a model describing the evolution of the state of the system over time (the transition equation)

$$x_t = f_t(x_{t-1}, v_{t-1}) \tag{2.1}$$

and a model relating the noisy measurements to the state (the measurement equation)

$$y_t = g_t(x_t, u_t) \tag{2.2}$$

[29] This section is summarized based on several survey papers including Javaheri, Lautier, and Galli (2003), Doucet, Godsill and Andrieu (2000), and Arulampalm, Maskell, Gordon, and Clapp (2002).

where u_t and v_t are mutually independent and identically distributed sequences with known probability density functions; $f(.)$ and $g(.)$ are possibly nonlinear but known functions; and the associated transition and measurement density, $p(x_t \mid x_{t-1})$ and $p(y_t \mid x_t)$ are determined by the density of u_t and v_t. In general, the hidden states can be either discrete-valued, continuous-valued or a mixture of the two. For exposition, we assume that the states are continuous.

All relevant information about the unobservable or hidden states $\{x_{0,\ldots,}x_t\}$ given observations up to and including time *t*, $\{y_{0,\ldots,}y_t\}$, can be obtained from the posterior distribution $p(x_{0:t} \mid y_{0:t})$. Estimating recursively in time two of its marginal densities is the main goal of optimal filtering. The first marginal is the filtering distribution

$$p(x_t \mid y_0,\ldots,y_t) = p(x_t \mid y_{0:t}) \tag{2.3}$$

which uses information up until time *t*. The second marginal, the smoothing distribution, uses all the information in the sample of size *T* to estimate past realizations of the state

$$p(x_t \mid y_0,\ldots,y_T) = p(x_t \mid y_{0:T}). \tag{2.4}$$

Obtaining the filtering density consists of essentially two steps: prediction and update. The prediction step computes the prior density of the state at time *t* based on *t-1* information using the Chapman-Kolmogorov equation

$$\begin{aligned} p(x_t \mid y_{1:t-1}) &= \int p(x_t \mid x_{t-1}, y_{1:t-1}) p(x_{t-1} \mid y_{1:t-1}) dx_{t-1} \\ &= \int p(x_t \mid x_{t-1}) p(x_{t-1} \mid y_{1:t-1}) dx_{t-1} \end{aligned} \tag{2.5}$$

where $p(x_t \mid x_{t-1})$ is the transition density. When the measurement y_t becomes available at time t, the update step modifies the prior to obtain the posterior density for the current state, via Bayes's rule

$$p(x_t \mid y_{1:t}) = \frac{p(y_t \mid x_t)p(x_t \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})} \tag{2.6}$$

with $p(y_t \mid y_{1:t-1}) = \int p(y_t \mid x_t)p(x_t \mid y_{1:t-1})dx_t$, where $p(y_t \mid x_t)$ is the measurement density.

The optimal filter, and therefore the solutions to the equation (2.5) and (2.6), is analytically available in only a few special cases. One is the Kalman filter, which computes the filtering distribution exactly when the functions $f(.)$ and $g(.)$ are linear and both u_t and v_t are Gaussian. The optimal filter can also be calculated when the state space is discrete and consists of a finite number of states. In this case, the model is known as a discrete hidden Markov model (HMM), see Cappe, et. al. (2005) for a review. [30] However, the assumptions made above may not hold in situations of interest; for example, many dynamic systems are non-linear, non-Gaussian, or are more appropriately modeled in continuous-time. Therefore, enormous efforts have been devoted to approximating these filtering distributions when it is impossible to evaluate them analytically. These approaches include the extended Kalman filter, approximate gird-based filters, and particle filters. We focus on particle filters in this paper, which have shown to be more flexible and reliable (see Javaheri, Lautier, and Galli (2003),

[30] Regime switching models as in Hamilton (1987) do not fall into the same category as hidden Markov models because they depend upon *past* latent discrete states. The theoretical properties and computational methods for the estimators are different.

Doucet, Godsill and Andrieu (2000), and Arulampalm, Maskell, Gordon, and Clapp (2002)).

I: Particle Filtering via Sequential Importance Resampling (SIR)

SIR is a commonly used particle filtering algorithm, which approximates the filtering distribution by a weighted set of particles.[31] The weights are chosen using the principle of *importance sampling*, which can be illustrated as follows. Suppose there exists a probability density of interest called the target density $\pi(x)$ from which it is difficult to draw samples but that can be evaluated up to a constant of proportionality $\tilde{\pi}(x)$. The main idea behind importance sampling is to draw randomly from a distribution other than $\pi(x)$ and reweight each draw so that it is approximately an i.i.d. realization from $\pi(x)$. Taking N_s random draws from a proposal density $q(x)$, called an *importance density*, the realized sample $\{x^i\}_{i=1}^{N_s}$ gets reweighted by calculating importance weights $w^i \propto \frac{\tilde{\pi}(x^i)}{q(x^i)}$ for each draw. The approximation to the density $\pi(x)$ is given by $\pi(x) \approx \sum_{i=1}^{N_s} w^i \delta(x - x^i)$ where $\delta(\cdot)$ denotes the Dirac mass. This weighted approximation of $\pi(x)$ approaches the true posterior density as $N_s \to \infty$.

Therefore, for our situation here, the set of points $\{x_t^i, w_t^i\}_{i=1}^{N_s}$ characterizes the posterior density $p(x_t \mid y_{1:t})$, where $\{x_t^i\}_{i=1}^{N_s}$ is a set of support points, called particles,

[31] See Doucet et. al. (2001) for a book length review of particle filtering.

with associated weights $\{w_t^i\}_{i=1}^{N_s}$ that are normalized such that $\sum_{i=1}^{N_s} w^i = 1$. At each iteration, the posterior density at t will be approximated as

$$p(x_t \mid y_{1:t}) \approx \sum_{i=1}^{N_s} w_t^i \delta(x_t - x_t^i) \tag{2.7}.$$

With the addition of a new observation, the next posterior density can be approximated by taking the particles simulated in the past and moving them to a new position on the support of the density. The new set of particles will then require a new set of importance weights to accurately approximate the new density. If the sample x_t^i was drawn from an importance density $q(x_t \mid y_{1:t})$, then the importance weights at any point in time are

$$w_t^i \propto \frac{p(x_t^i \mid y_{1:t})}{q(x_t^i \mid y_{1:t})} \tag{2.8}$$

In practice, the importance weights can be updated recursively. The denominator in (2.8) can be rewritten as

$$q(x_t^i \mid y_{1:t}) = q(x_t^i \mid x_{t-1}^i, y_{1:t}) q(x_{t-1}^i \mid y_{1:t-1}) \tag{2.9};$$

while the numerator can be rewritten as

$$\begin{aligned} p(x_t^i \mid y_{1:t}) &= \frac{p(y_t \mid x_t^i, y_{1:t-1}) p(x_t^i \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})} \\ &= \frac{p(y_t \mid x_t^i, y_{1:t-1}) p(x_t^i \mid x_{t-1}^i, y_{1:t-1}) p(x_{t-1}^i \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})} \\ &\propto p(y_t \mid x_t^i) p(x_t^i \mid x_{t-1}^i) p(x_{t-1}^i \mid y_{1:t-1}) \end{aligned} \tag{2.10}.$$

Substituting (2.9) and (2.10) into (2.8), the equation for updating the importance weights recursively is

$$w_t^i \propto \frac{p(y_t \mid x_t^i)p(x_t^i \mid x_{t-1}^i)p(x_{t-1}^i \mid y_{1:t-1})}{q(x_t^i \mid x_{t-1}^i, y_{1:t})q(x_{t-1}^i \mid y_{1:t-1})}$$
$$= w_{t-1}^i \frac{p(y_t \mid x_t^i)p(x_t^i \mid x_{t-1}^i)}{q(x_t^i \mid x_{t-1}^i, y_{1:t})} \qquad (2.13).$$

One major issue with this algorithm is that the variance of the importance weights increases randomly over time. After a few iterations of the algorithm, the majority of the probability mass will be allocated to only a few particles and the collection of particles will be a poor representation of the density. This phenomenon has been labeled *degeneracy* in the literature; see Doucet et al. (2001). In order to solve this problem, a resampling step gets added to the algorithm which maps the unequally weighted set of particles $\{x_t^i\}_{i=1}^{N_s}$ to a new set of equally weighted particles. The purpose of resampling is to make copies of the "better" particles at each iteration in order to explore the support of the next posterior density. The simplest resampling algorithm is known as multinomial resampling, which corresponds to drawing new particles from a multinomial distribution based on the set of normalized importance weights $\{w_t^i\}_{i=1}^{N_s}$. This algorithm can be improved upon using stratification methods as proposed in Carpenter, Clifford, and Fearnhead (1999). In addition, resampling does not need to be performed at every iteration but instead should be conducted randomly when the variance of the importance weights grows. Liu et al. (1998) introduced a measure called the *effective sample size* (ESS), defined as $ESS = 1/\sum_{i=1}^{N_s}(w_t^i)^2$, to determine when resampling should be performed.

The most important part of any particle filtering algorithm is the choice of the importance density $q(x_t \mid x_{t-1}^i, y_{1:t})$. Ideally, we would like to choose an importance density that approximates the target density as closely as possible. The particle filtering literature includes the concept of a conditionally optimal importance density and many other suboptimal approximations using local linearization techniques are available; see Doucet et al. (2000). A convenient choice used by many authors is the prior density

$$q(x_t \mid x_{t-1}^i, y_{1:t}) = p(x_t \mid x_{t-1}^i) \tag{2.14}$$

because substituting (2.14) into (2.13) leads to a particularly simple method to update the importance weights

$$w_t^i = w_{t-1}^i p(y_t \mid x_t^i) \tag{2.15}.$$

Therefore, a generic SIR algorithm is described by

For $t = 1,...,T$

Step 1:

For $i = 1,...,N_s$

Draw $x_t^i \sim p(x_t \mid x_{t-1}^i)$ and compute the importance weights w_t^i using (2.15)

End For

Step 2:

For $i = 1,...,N_s$

Normalize the importance weights as $w_t^i = w_t^i / \sum_{i=1}^{N_s} w_t^i$

End For

Step 3:

Resample the particles using normalized weights if the ESS falls below a specified threshold; that is, multiply/discard particles w.r.t. high/low importance weights to obtain particles with the weights set equal.

End For

II: Particle Filtering via Auxiliary Particle Filters (APF)

The SIS algorithm (the SIR algorithm without the resampling step) provides the basis for most particle filters that have been developed so far. Various versions of particle filters proposed in the literature can be regarded as special cases derived from the general SIR algorithm by an appropriate choice of the importance density and/or modification of the resampling step. More recently, Pitt and Shephard (1999) suggested the use of auxiliary particle filters (APF) to overcome some difficulties in the SIS/SIR algorithm.

The idea of the APF algorithm is to obtain a sample from the joint density $p(x_t, i \mid y_{1:t})$ and then omitting the indices i in the pair (x_t, i) to produce a sample $\left\{x_t^j\right\}_{j=1}^{N_s}$ from the marginalized density $p(x_t \mid y_{1:t})$. The importance density $q(x_t, i \mid y_{1:t})$ is used to draw the sample $\left\{x_t^j, i^j\right\}_{j=1}^{N_s}$, where i^j refers to the index of the particles at $t-1$. Therefore, the resampling weights could be constructed as $w_t^j \propto \frac{p(x_t^j, i^j \mid y_{1:t})}{q(x_t^j, i^j \mid y_{1:t})}$. Specifically, a proportionality can be derived for $p(x_t, i \mid y_{1:t})$ as

$$p(x_t, i \mid y_{1:t}) \propto p(y_t \mid x_t) p(x_t \mid x_{t-1}^i) w_{t-1}^i \tag{2.16}$$

and (2.16) can be approximated by

$$q(x_t, i \mid y_{1:t}) \propto p(y_t \mid \mu_t^i) p(x_t \mid x_{t-1}^i) w_{t-1}^i \tag{2.17}$$

where μ_t^i is the mean, the mode, a draw, or some other likely values associated with the density of $x_t \mid x_{t-1}^i$. The form of the approximating density is designed so that

$$q(i \mid y_{1:t}) \propto p(y_t \mid \mu_t^i) w_{t-1}^i \tag{2.18}$$

Then the weight assigned to the sample $\left\{x_t^j, i^j\right\}_{j=1}^{N_s}$ is given by

$$w_t^j \propto w_{t-1}^{i^j} \frac{p(y_t \mid x_t^j) p(x_t^j \mid x_{t-1}^{i^j})}{q(x_t^j, i^j \mid y_{1:t})} = \frac{p(y_t \mid x_t^j)}{p(y_t \mid \mu_t^{i^j})} \tag{2.19}$$

Thus, we can sample from $q(x_t, i \mid y_{1:t})$ by simulating the index with probability $w_t^j \propto q(i \mid y_{1:t})$ and then sampling from the transition density $p(x_t \mid x_{t-1}^i)$. The w_t^i's are called the first-stage weights. Having sampled the joint density of $q(x_t, i \mid y_{1:t})$ N_s times, we perform a reweighting, putting on the draw $\left\{x_t^j, i^j\right\}_{j=1}^{N_s}$ the weights proportional to the so-called second-stage weights defined in (2.19) with the hope that these second-stage weights are much less variable than for the original SIR method.

The APF algorithm is described by

For $t = 1, \ldots, T$

Auxiliary variable resampling step:

Step 1:

For $i = 1, \ldots, N_s$

Compute μ_t^i

Compute the first-stage importance weights $w_t^i = q(i \mid y_{1:t}) \propto p(y_t \mid \mu_t^i) w_{t-1}^i$

End For

Step 2:

For $i = 1, \ldots, N_s$

Normalize the first-stage importance weights as $w_t^i = w_t^i / \sum_{i=1}^{N_s} w_t^i$

End For

Step 3:

Resample the particles using normalized weights; that is, multiply/discard particles w.r.t. high/low importance weights to obtain particles with the weights set equal.

End For

Importance Sampling Step:

Step 4:

For $j = 1, \ldots, N_s$

Draw $x_t^j \sim q(x_t \mid i^j, y_{1:t-1}) = p(x_t \mid x_{t-1}^{i^j})$ as in the SIR filter

Compute the second-stage importance weights w_t^j using (2.19)

End For

Step 5:

For $i = 1, \ldots, N_s$

Normalize the second-stage importance weights as $w_t^j = w_t^j / \sum_{j=1}^{N_s} w_t^j$

End For

End For

III. Smoothing via Particle Methods

Calculation of the marginal smoothing density $p(x_t \mid y_{0:T})$ is a separate challenge in non-linear, non-Gaussian state space models. The filtering literature has demonstrated that the smoothing density can be calculated recursively in two different ways: two-filter formula smoothing and backwards Markovian smoothing, see Cappe (2005) for details. Both methods were developed in the 1960-1970's for the linear Gaussian state space model and their counterparts exist for particle methods. The first particle smoothing algorithm based upon two-filter formula smoothing was proposed by Kitagawa (1996). Doucet et. al. (2000) developed a backwards Markovian particle smoother that is particularly simple to implement and we apply it in this paper. Particle smoothing algorithms based on the backward Markovian algorithm can be computationally time intensive, although recent developments by Klass et. al. (2006) aim to dramatically decrease this time.

The Backwards Markovian Smoothing algorithm is described by

At $t = T$

For $i = 1,...,N_s$ set $\tilde{w}_{t|t}^i = w_t^i$

For $t = n-1,.....,0$

For $i = 1,...,N_s$ evaluate the importance weight

$$\tilde{w}_{t|n}^{i} = \sum_{j=1}^{N_s} \tilde{w}_{t+1|n}^{j} \frac{w_t^i p\left(x_{t+1}{}^{j} \mid x_t^{i}\right)}{\left[\sum_{l=1}^{N_s} w_t^l p\left(x_{t+1}{}^{j} \mid x_t^{l}\right)\right]}$$

2.3. Models and EMM Estimations

In this section, the EMM methodology is briefly described, details of which are presented in Chapter 1, Section 2. The EMM estimation results for fitting the US short rate are summarized for two continuous-time SV models (a two-factor and a three-factor SV models) and two discrete-time RS models (a RS-in-volatility model with and without the level effect), which have passed the EMM specification tests and well fitted the US short rate.

2.3.1. EMM Methodology

The theory of EMM estimation is developed in Gallant and Tauchen (1996) and is extended to non-Markovian data with latent variables in Gallant and Long (1997). The basic procedure of EMM estimation consists of two steps. First, the empirical conditional density of observed time series is estimated by a seminonparametric (SNP) series expansion. This SNP expansion has a VAR-(G)ARCH Gaussian density as its leading term, and departures from the Gaussian leading term are captured by a Hermite polynomial expansion. This step is accomplished by projecting the data onto a SNP model and is therefore called the projection step. Second, a GMM-type criterion function is constructed using the score functions (from the log-likelihood of the SNP

density) as moments. The scores are evaluated using the simulation output from a given structural model and the criterion function is minimized with respect to the parameters underlying the structural model. This step is to extract structural parameters from the data summary by minimizing the χ^2 criterion and is therefore called the estimation step. The expository discussion of the method is in Gallant and Tauchen (2001). A summarized detail of the ideas refers to Chapter 1.

2.3.2. Data

Weekly (Wednesday) observations of the annualized yield on the 3-month US T-bill (over the period January 1954 to September 2004) are used for the empirical work.[32] The raw data plotted in Figure 1.2, and descriptive statistics are given in Table 1.2. The basic stylized facts concerning the short-rate are: near non-stationary behavior (slow mean reversion), large changes and small changes are clustered together (ARCH effect), the volatility of rates increases with the level of rates (level effect), and positive skewness and excess kurtosis (non Gaussian distribution). The non Gaussian behavior of the short rate is clearly shown by the qq-plot in Figure 1.4 and by the statistics summary in Table 1.2; the slow mean reversion and ARCH effect are vividly illustrated in the autocorrelation plots in Figure 1.3.

2.3.3. SV Models and EMM Estimations

Two candidate SV models are the following continuous-time diffusion models:

---

[32] See more details about the data source and data selection in Chapter 1, Section 4.

SV2 Model:
$$\begin{cases} dr_t = (\phi_0 - \phi_1 r_t)dt + r_t^{\gamma}\sigma_t dz_t = k_r(\mu_r - r_t)dt + r_t^{\gamma}\sigma_t dW_1 \\ d\log(\sigma_t^2) = (\omega_0 + \omega_1 \log(\sigma_t^2))dt + \xi dW_2 \end{cases} \tag{2.20}$$

SV3 Model:
$$\begin{cases} dr_t = (\phi_0 - \phi_1 r_t)dt + \sigma_t r_t^{\gamma} dz_t = k_r(\mu_{r,t} - r_t)dt + \sigma_t r_t^{\gamma} dW_1 \\ d\log(\sigma_t^2) = (\omega_0 + \omega_1 \log(\sigma_t^2))dt + \xi dW_2 \\ d\mu_t = (\upsilon_0 + \upsilon_1 \mu_t)dt + \zeta dW_3 \end{cases} \tag{2.21}$$

where $\{r_t\}$ is the short rate at time t, and dW_1, dW_2 and dW_3 are mutually independent standard Brownian motions.

The two-factor and three-factor SV models (hereafter, the SV2 and SV3 model) are extended from the CKLS model, given by

$$dr_t = (\phi_0 - \phi_1 r_t)dt + r_t^{\gamma}\sigma dz_t = k_r(\mu_r - r_t)dt + r_t^{\gamma}\sigma dW \tag{2.22}$$

First presented by Chan et al. (1992) as a generalized diffusion model[33], the key characteristic of the CKLS model is that the conditional mean and variance of changes of the series depend on the level of the series. Specifically, r_t mean-reverts towards the long-run level μ_r, with the speed of the reversion, k_r, and γ captures the influence of the level of series to the conditional volatility (also called the *level effect*). In the spirit of Taylor (1986, 1994), our two SV models introduce additional latent factors on top of the CKLS model. The SV2 model allows the log-volatility of short rate to follow a mean reverting process as well as the series itself. Moreover, the SV3 model introduces another factor associated with the reverting mean level, that is, the long-run mean is assumed to follow mean-reverting process as well as the log-volatility of short rate

[33] Many well known models can be nested with appropriated parameter restrictions within the CKLS model.

series. The sensitivity of shocks to the log-volatility and to the long-run mean are measured by the non-negative parameters ξ and ζ, respectively.[34]

The preferred SNP model for describing the short rate data (a AR(1)-GARCH(1,1)-*Kz*(7) SNP model) and the EMM estimation results for the above two SV models are reported in Chapter 1, Section 5. Here, we briefly outline the major findings. Firstly, the small *p*-value based on the χ^2 distribution associated with the EMM statistic, led to a strong rejection of the CKLS model. On the other hand, the SV2 model and the SV3 model were not rejected at the 10% level, implying the introduction of additional stochastic factor(s) is important for explaining the behavior of US short rate. The SV3 model was even favored over the SV3 model; that is, the fitting improved further when adding a stochastic mean factor. Secondly, strong mean reverting behaviors were confirmed for both models with similar reverting trend, measured by ϕ_0 / ϕ_1 or υ_0 / υ_1, around 2.50%. This estimate is lower than the previous works, partially due to the incorporation of longer data series after 1989. In addition, the implied log-volatility process for both models and the reverting mean process for the SV3 model are highly persistent. Thirdly, the conditional volatility is sensitive to the level of the rates for both models; the level effect estimates, γ, are significantly in excess of zero. Specifically, γ is at 0.66 and 0.51 for the SV2 and SV3 model, respectively; both estimates are well below unity. Lastly, adjusted *t*-ratios of all individual elements of the score vector are well below 2.0 for the accepted SV2 and SV3 model, implying

[34] See detailed discussion in Chapter 1, section 3.1.A.

particularly that both model have no difficulties to capture the volatility clustering that exists in the data as summarized by the 11117000 SNP model.

2.3.4. RS Models and EMM Estimations

We also consider two successful RS candidates as follows:

RS-in-σ model: $$\Delta r_t = \phi_0 - \phi_1 r_{t-1} + \sigma_i z_t \tag{2.23}$$

RS-in-σ + *Level* model: $$\Delta r_t = \phi_0 - \phi_1 r_{t-1} + \sigma_i r_{t-1}^{\gamma} z_t \tag{2.24}$$

where the regime indicator is $i = 1, 2$. The switching states are governed by a first-order Markov process with the matrix of transition probabilities as

$$P = \begin{pmatrix} P_{11} & 1 - P_{11} \\ 1 - P_{22} & P_{22} \end{pmatrix} \tag{2.25}$$

where $p_{ij} = \Pr(S_t = j \mid S_{t-1} = i)$ is the transition probabilities from regime j to regime i.

The simple RS-in-σ model is a RS version of the discrete-time OU process: the mean dynamics are assumed to be mean reverting, but regime-switching behavior is allowed for the conditional variance dynamics. That is, volatility shocks are regime-dependent in order to accommodate time-varying volatility. Built upon the generalized CKLS process, the RS-in-σ + *Level* model incorporates the level effect as well as the RS-in-volatility effect to accommodate the time-varying behavior and conditional heteroskedasticity. Specifically, the sensitivity of volatility to the level of the rate is measured by γ, and the conditional volatility switches between two very persistent regimes.

Using the preferred SNP model 11117000, the EMM estimation results for the

above two RS models are provided in Table 1.7c. Some major results are briefly summarized as follows. Firstly, the p-value for the RS-in-σ model is slightly higher than 5%, providing mild evidence in support of the model; the RS-in-σ + *Level* model, with a *p*-value of 0.29, shows significant improvement over the RS model without incorporating the level effect. It fits even better than the SV3 model, in which both the level effect and SV effect are implemented in the underlying structural model. Secondly, strong mean reversions are confirmed for both RS models with the long-run reverting means estimated higher than those of the SV models. Thirdly, a low-volatility and a high-volatility regime are identified with highly persistent transition probabilities; both probabilities exceed 0.90 for two RS models. The transition probability of staying in the high-volatility regime is estimated smaller than the previous works, for example, Smith (2002). Lastly, the estimated conditional volatility in the RS-in-σ + *Level* model is sensitive to the level of the short rates. The magnitude of the level effect is much lower than unity but very similar to that was found in our multi-factor SV models.

2.4. Empirical Results

In this section, we first demonstrate the performance of the particle filter methodology by applying the SIR and APF algorithms on data that are simulated from the SV2 and SV3 models using the EMM estimation results. Then we apply the particle filter algorithms to the US short-term interest rate and analyze the filtering and smoothing results in section 2.4.1. Similar procedures are processed for the two RS models using the HMM algorithm; the results are shown in section 2.4.2.

2.4.1 Results for SV Models

To show the performance of the particle filter algorithms, we simulated representative paths from the continuous-time SV2 and SV3 model using the Euler scheme to approximate the solutions to the SDEs of equation (2.1) and (2.2). In order to mimic the real process with a weekly frequency, 2500 simulated values of the short-term interest rate (as well as its log-volatility and/or reverting mean process) are separated by an implied time length of 1/52 (one week) because the EMM estimated parameters are assumed to generate annualized data, and successive values are separated by 25 internal discretization steps. With a burn-in period of 500 simulation steps, that is a total of (2500+500)*25=75000 discretization steps of size 1/(52*25) =0.008 for the entire simulation. The EMM estimates used for simulating the SV2 model and the SV3 model are provided in Table 1.7a.

Using the above simulation assumptions and EMM estimates, the plots of the simulated series are provided in the top panels of Figure 2.1 and Figure 2.3, respectively; the descriptive summary is shown in Table 2.1. Although the simulations from the fitted SV models are not exactly the same as the actual historical paths of the interest rate, they are capable of generating extreme volatile periods as the monetary experiment experience during 1979-1982. They also share qualitative features of the real data as the positive skewness and high excess kurtosis. The true but latent processes of the log-volatility process, volatility process, and/or the reverting mean process are also presented in Figures 2.1 and 2.3.

Our task is to apply the SIR algorithm and APF algorithms to the *observed* interest rate data as simulated above to filter out the latent processes as accurately as

possible. With the simulated data, we can compare the filtered estimates to the true simulated latent processes, which is sometimes what we cannot do with actual interest rate data. Notice that it is beyond the scope of this paper to explore how discretization bias, sampling frequency, or parameter estimation affects the accuracy of volatility filtering estimates; so we just focus on the performance of particle filters using the current simulation assumptions that are consistent with the EMM estimates for the weekly US short rate.

Figures 2.2 and 2.4 provide a graphical depiction of the performance of the two particle filtering algorithms on the SV2 and SV3 model. The top panels display the filtered and smoothed latent process(es) relative to the true process(es) using the SIR algorithm; the bottom panels are using the APF algorithm. Table 2.2 reports the associated statistics of root mean squared error (RMSE), which indicates the degree of the deviations of the filtered/smoothed series from the true series. A number of noticeable results are as follows. First, the particle filter methodology performs extremely well on the simulated data. For example, the estimated log-volatility process closely captures the true latent process as clearly shown in Figure 2.2. The small RMSEs imply that the estimation errors are very well controlled. Secondly, the results slightly improve from the filtering step using the information up to the current state to the smoothing step using the complete information, which is true for both SIR and APF algorithm. Thirdly, there is no significant advantage of using one algorithm over the other in terms of the estimation error for the log-volatility process; the SIR and APF algorithms result in close RMSEs to each other at each estimation step. Lastly, the estimate results for the mean process in the SV3 models are a bit worse than these for

the volatility process. There is no improvement of the estimation accuracy from the filtering step to smoothing step; the RMSE is actually a bit larger using either SIR or APF algorithm. In addition, both of the filtered and smoothed series for the mean process appear to have some problem to capture the highly volatile parts of the true mean process.

We now apply the SIR/APF algorithm to the US short rate series: Figure 2.5 provides the estimation results for the volatility process in the SV2 model; Figures 2.6a and 2.6b are for the volatility process and mean process in the SV3 model, respectively. Each figure displays the plots of the series of the short rate with its first-order difference, and the filtered and the smoothed series using the two particle filter algorithms. For the SV2 model, the estimations for the latent volatility process from the filtering step and the smoothing step follow similar paths closely, and the results are very similar using different particle filter algorithms. Moreover, a number of high-volatility periods that were implied from the filtering/smoothing estimations can well resemble the historical business cycles and important events during the data period of 1954 to 2004. Specifically, the volatility process has two spikes in late 1950's and early 1960's, corresponding to the two short recessions during 8/1957-4/1958 and 4/1960-2/1961. After the second spike, the volatilities decrease and keep historically low until the mid 1960's. Then a high-volatility period is experienced for the short rate during the 1970's, in which occurred the stagflation in the early 1970s when recession and inflation occurred simultaneous, the energy crisis from 1973 to 1975 due to the onset of an oil embargo by OPEC, and the monetary experiment that was conducted by the Federal Reserve during 1979-82 when its policy shifted away from targeting the federal funds

rate. The volatility process gets to a relatively stable period since the mid 1980's. Two significant jumps of the volatilities are due to the stock market crash of October 19, 1987 and the terrorist attack on September 11, 2001.

Although the SV3 model fits the historical data better than the SV2 model, the filtered/smoothed volatility process in the SV3 model shares a similar stochastic path with that in the SV2 model. The estimated volatilities from the SV3 model during the high-volatility periods appear a bit higher than those from the SV2 model; for example, during the monetary experiment period and the market crash of October 19, 1987. For the mean process of the SV3 model, the estimations from the filtering and the smoothing step slightly deviate from each other. Specifically, the filtered estimation is consistently lower than the smoothed estimation for most of the time for both SIR/APF algorithms. Moreover, a noticeable finding about the filtered/smoothed mean process is that the reverting mean is relatively high when the volatility is low but is relatively low when the volatility is high. For example, the US economy experienced an unprecedented period of long economic growth and a bull stock market run since the mid 1990's. During this period, the volatilities of the short rate were enduringly low but the reverting means kept increasing. When this lasting peacetime economic growth was ended by the terrorist attack on September 11, 2001, the short rate became more volatile but its reverting means dropped significantly.

2.4.2 Filtering Results for RS Models

To validate the HMM filtering and smoothing algorithm for the RS model, we first simulated representative paths from the discrete-time RS-in-σ model and RS-in-

$\sigma + Level$ model described by equation (2.1) and (2.2). 2500 observations are simulated using the function `simulation.MSAR` in `S+FinMetrics 2.0`, and the series of the switching indicator i's is saved. The EMM estimates used for simulating the two RS models are provided in Table 1.7c. The simulations are plotted in the top panels of Figure 2.7 and Figure 2.8; the descriptive summary is shown in Table 2.1. Unlike the simulations from the fitted SV models, those from the RS models appear to generate higher unconditional means and lower excess kurtosis values. In addition, the regime-switching indicators are also presented in Figures 2.7 and 2.8 with 0/1 as the low/high-volatility regime.

The bottom panels in Figures 2.7 and 2.8 provide a graphical depiction of the estimation performance for two RS models using the HMM filter and smoother as described in Cappe et. al. (2005). The third panel displays the true process of the switching indicator while the 4^{th}-6^{th} panels plot the filtered and smoothed probabilities as well as the joint estimate of the overall sequence of states. As a measurement of the estimation error, the corresponding RMSEs are reported in Table 2.2. Some findings on the estimation performance for both RS models are as follows. (1) The estimation using the information up to the current state is very noisy. Clearly shown by the Figure 2.7, the plots of the estimated filtered probabilities are very choppy. (2) The estimations significantly improve when the information set expands to involve all the states; the associated RMSEs for the smoothed series decrease greatly. (3) In a discrete setting, the smoothing algorithm computes the probability of being in a state at a given point in time but this does not take into account the *joint* behavior of other states. Therefore, we also implemented the Viterbi algorithm as described in Cappe et. al. (2005) for this set

of models, which computes the maximum a posterior (MAP) estimator of the sequence of state variables. That is, the Viterbi algorithm is a dynamic programming algorithm that computes the most likely *joint* sequence of states that occurred through time given the complete information set. This algorithm allows us to answer a fundamentally different question: given the data, which sequence of states was most likely? Compared with the true process of the switching indicator, the estimated smoothed probabilities and the joint sequence of states are capable of capturing most of the regime switches.

Lastly, we apply the HMM algorithm to the US short rate series and provide the estimation results for the two RS models in Figures 2.9 and 2.10, respectively. The panels display the plots for the series of the US short rate with its first-order difference, the filtered and smoothed probabilities from the HMM algorithm, and finally the sequence of joint states from the Viterbi algorithm. Although the RS-in-$\sigma + Level$ model fits the historical time series much better than the RS-in-σ model, our filtering/smoothing results provide evidence that the RS-in-σ model can better resemble the historical business cycles and important events that may cause strong shifts in the behavior of interest rate dynamics. For both the RS-in-σ model and the RS-in-$\sigma + Level$ model, the estimation results are similar to what we have seen for the simulated data: the estimated switching probability is noisy using the information up to the current state, but improves greatly using all the information; see the smoothed probabilities and the joint sequence of states shown in the bottom panels of Figure 2.9 and 2.10. The joint states imply that there are twelve high-volatility regimes during 1954-2004. Specifically, the second and third high-volatility regimes, indicated from the graph, correspond to the two short recessions during 8/1957-4/1958 and 4/1960-

2/1961; the following three high-volatility regimes appear in the early 1970s when recession and inflation occurred simultaneous, and the seventh regime may result from the 1973 energy crisis. The eighth and ninth regimes correspond to the period of the monetary experiment that was conducted by the Federal Reserve during 1979-82. After that, the US economy experienced a long low-volatility period in the 1990's, which was an unprecedented period of long economic growth. Our estimation successfully captures two abnormal events that interrupted this lasting economic growth: one is the stock market crash of October 19, 1987; the other is the terrorist attack on September 11, 2001.

Figure 2.10 depicts the filtered and smoothed probabilities as well as the joint sequence of states for the RS-in-σ + *Level* model that accounts for the level effect of the interest rate as well as the regime-switching effect in the volatility. From this model, more high-volatility regimes have been identified and the estimated indicator switches regime much more frequently than that from the simpler RS-in-σ model. Consequently, the estimated switching probability is still noisy even using all the available information. In addition, the estimated switching probabilities are in less agreement with the NBER reference cycle and important historical events than the RS-in-σ model, although the RS-in-σ + *Level* model provided the highest p-value based on the EMM estimation and thus the best fitting performance across all the considered volatility models. Specifically, the RS-in-σ + *Level* model indicates that the interest rate stays in the high-volatility regime at most of the time during 1954-1963, but the change of the real interest rate is relatively low during that period; two volatile periods, corresponding to the short recessions late of 1950's and early of 1960's, are not clearly identified. Moreover,

besides the two high-volatility regimes, 1973-1975 (the energy crisis period) and 1979-1980 (the monetary experiment period), which could be easily observed from the time series of the change of the data, several additional high-volatility regimes are also identified, although they last shortly. Also, the well-known period of the monetary experiment during 1979-1982 appears to expand to 1985 according to the estimated regime obtained from this model. In addition, besides the two that correspond to the stock market crash and the terrorist attack, additional high-volatility regimes are also implied during 1990's, although the US economy and the short rate are relatively stable at that period. Based on all the above evidence, our estimation results appear to show that the RS-in-σ + *Level* model over-fits the historical data.

2.5. Conclusion

Recovering the latent volatility through filtering and smoothing is important for forecasting future volatility and providing complementary information on model selection. While there is a sizable literature that uses EMM methodology to estimate the volatility models, the issue of volatility tracking and filtering after the estimation has not been widely explored within the EMM literature. In this paper, we comparatively evaluate the performance of a variety of volatility models in terms of the filtered and smoothed estimates of the latent processes using appropriate algorithms. With the EMM estimates for the preferred candidate volatility models for the US short rate, we first demonstrate the reliability of algorithms, two particle filtering algorithms for the continuous-time SV models and the HMM filter for the discrete-time RS models, by applying them on simulated data. Then we applying them to models for the U.S. short

rate fitted by EMM and evaluate their abilities to track and filter volatility in different volatility models. For the continuous-time SV models, we find that the SIR and APF algorithm perform similarly for filtering and smoothing the latent processes. The estimated volatility processes from the SV2 and SV3 models follow similar dynamic paths, although the SV3 model fits the historical data better than the SV2 model. In addition, the filtered/smoothed volatility processes from either model can resemble the historical business cycles and important events during 1954 to 2004. We also notice from the SV3 model that the reverting means usually move in an opposite direction with the volatilities. For the discrete-time RS models, the estimated high-volatility regimes from the RS-in-σ model are similar to the high-volatility periods in the SV2/SV3 model. We also find that the estimated switching probabilities from the RS-in-σ + *Level* model are in less agreement with the historical business cycles and important historical events than from the RS-in-σ model, although the former model provided the best fitting performance across all the considered volatility models based on the EMM estimation. Our results indicate an over-fit of the time series for the RS-in-σ + *Level* model.

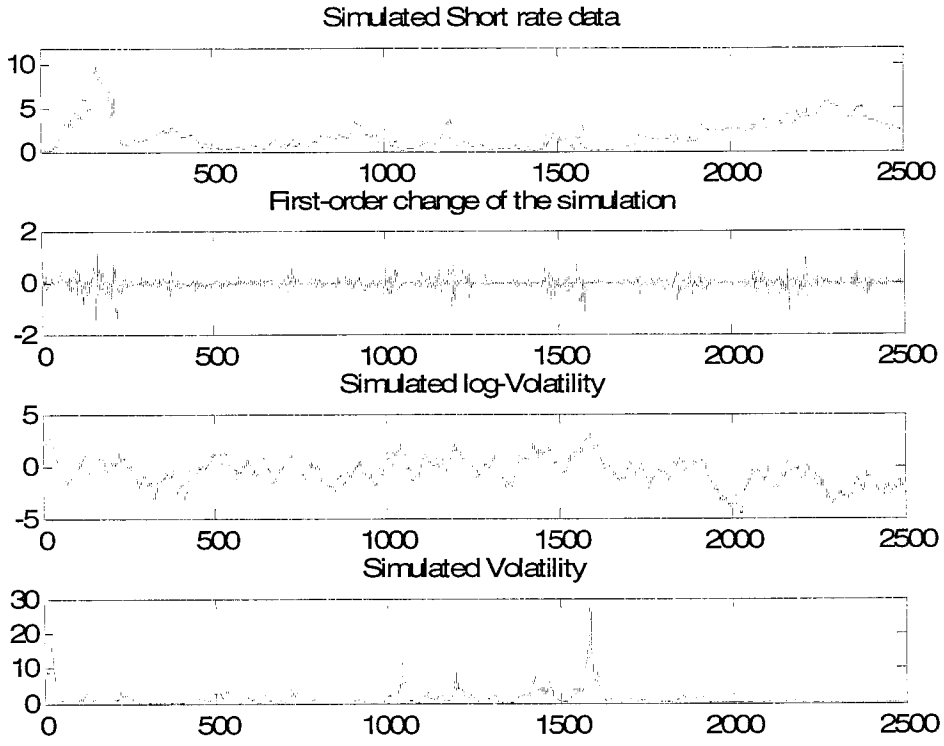
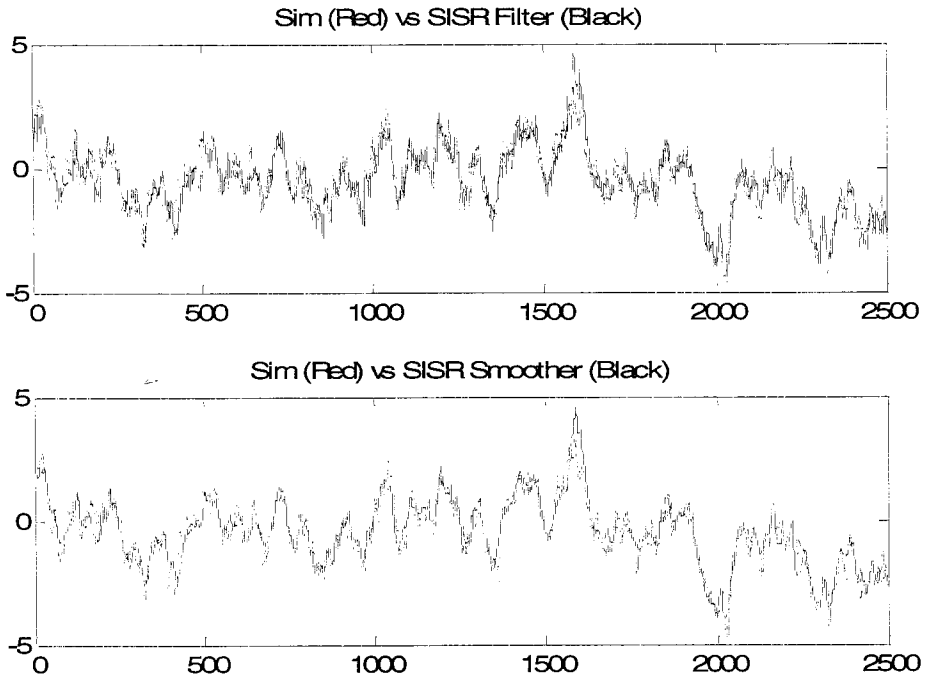


Figure 2.1: Simulated Series from SV2 Model

2500 observations are simulated from the two-factor SV model (2.1) with EMM estimates as shown in Figure 1.7. The simulated short rate series, its associated first-order difference, the latent log-volatility process, and the volatility process are plotted.

(A) SIR



(B) APF

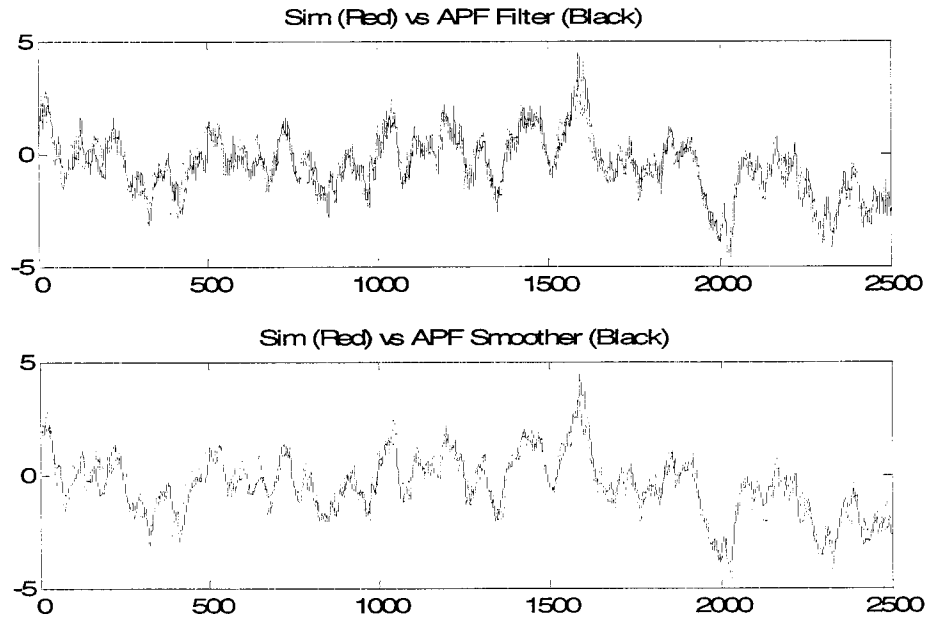


Figure 2.2: SIR/APF Results on Simulations from SV2 Model

Along with the *true* simulated log-volatility series (red lines), the corresponding estimation of filtered and smoothed series using the particle filter methods are plotted. The results using the SIR algorithm and the APF algorithm are presented in panel (A) and panel (B), respectively.

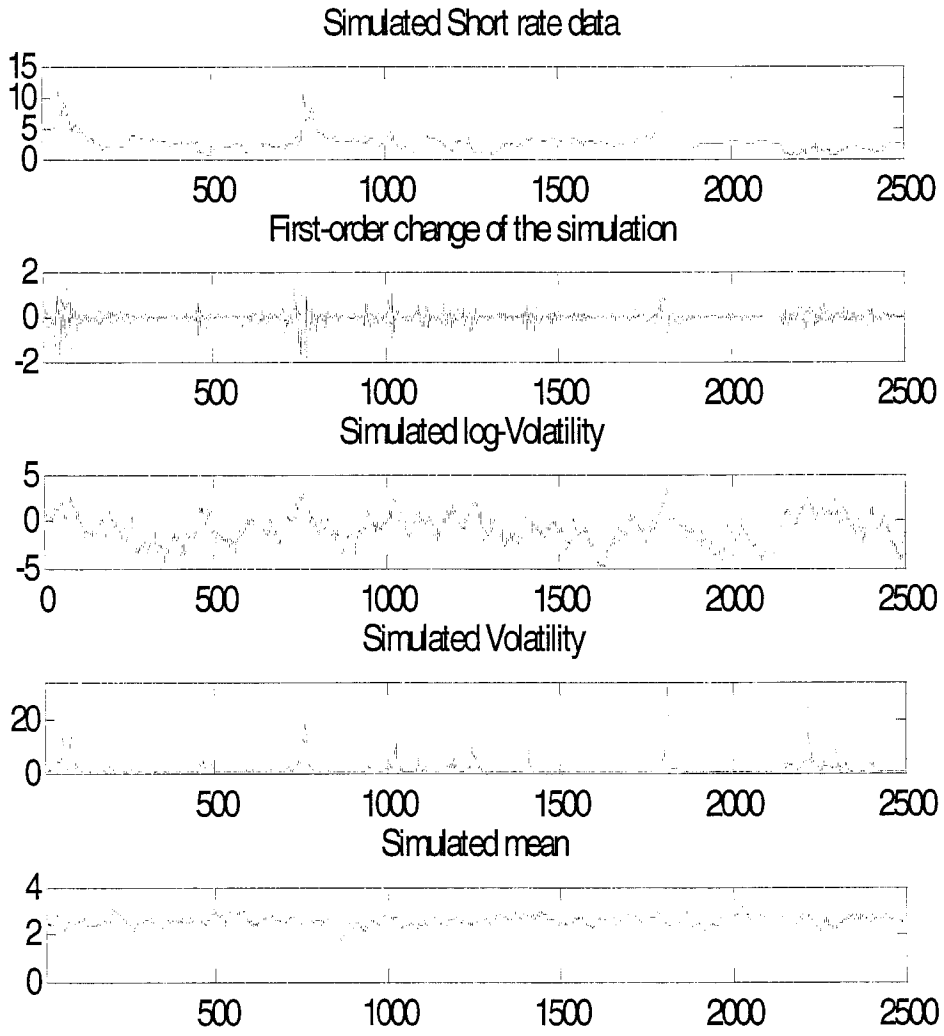
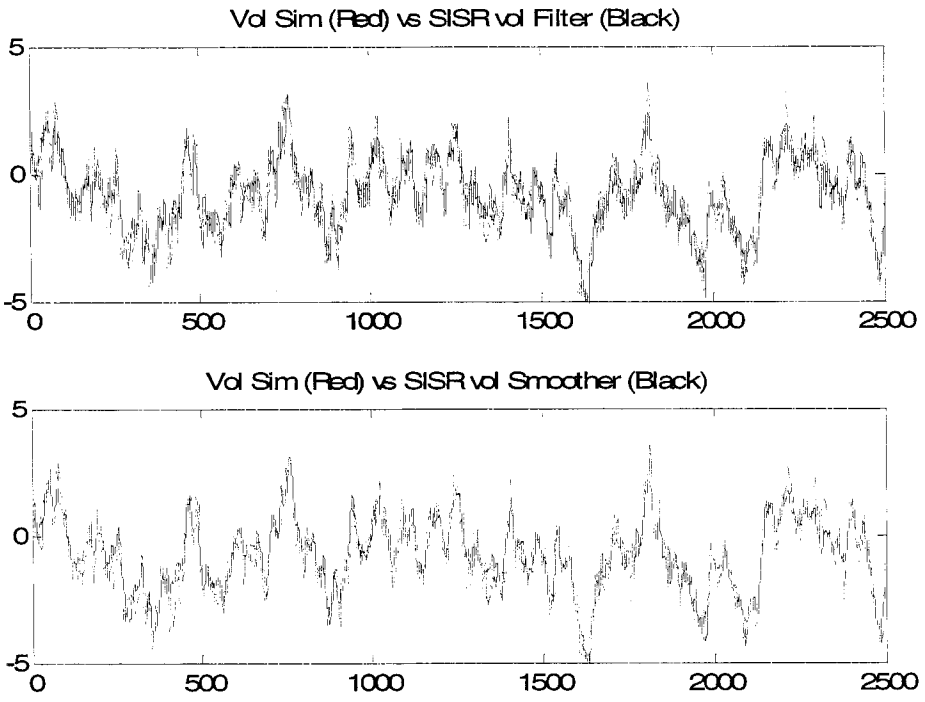


Figure 2.3: Simulated Series from SV3 Model

2500 observations are simulated from the three-factor SV model (2.2) with EMM estimates as shown in Figure 1.7. The simulated short rate series, its first-order difference, the latent log-volatility process, the volatility process, and the latent mean process are plotted.

(A) SIR



(B) APF

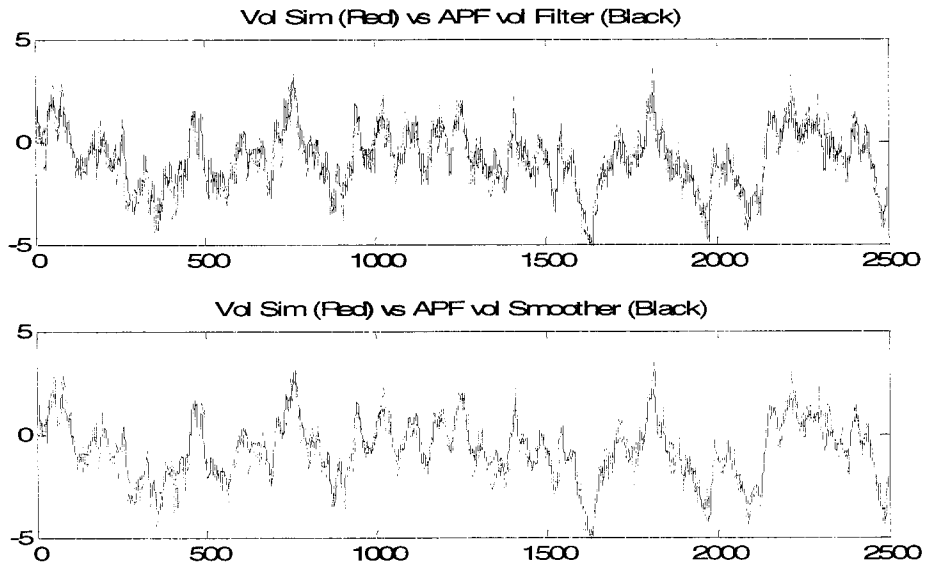
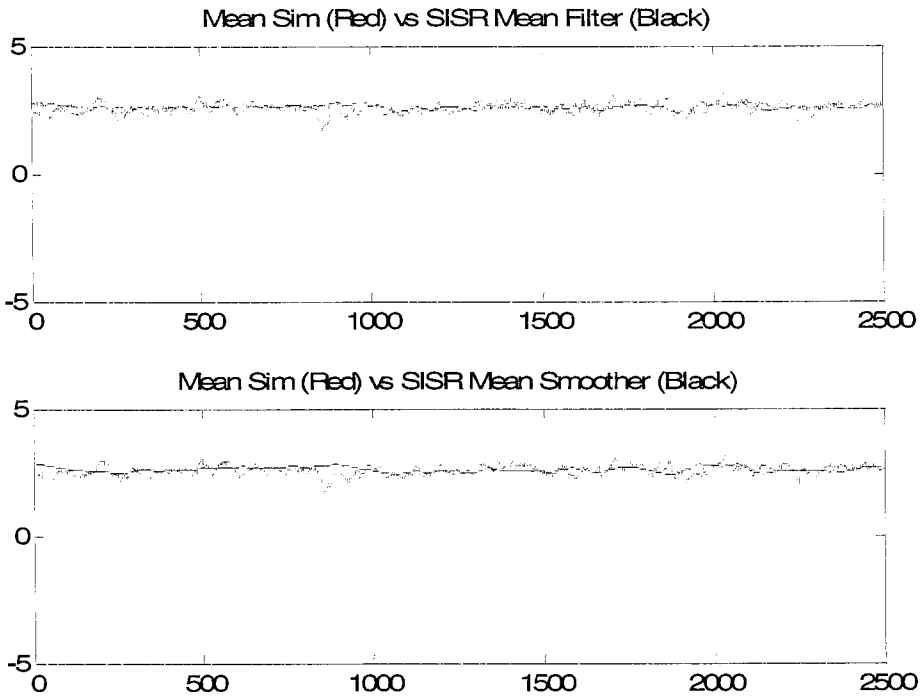


Figure 2.4a: SIR/APF Results on Simulations from SV3 Model

Along with the simulated log-volatility series (red lines), the corresponding estimation of filtered and smoothed series are plotted. The results using the SIR algorithm and the APF algorithm are presented in panel (A) and panel (B), respectively.

(A) SIR



(B) APF

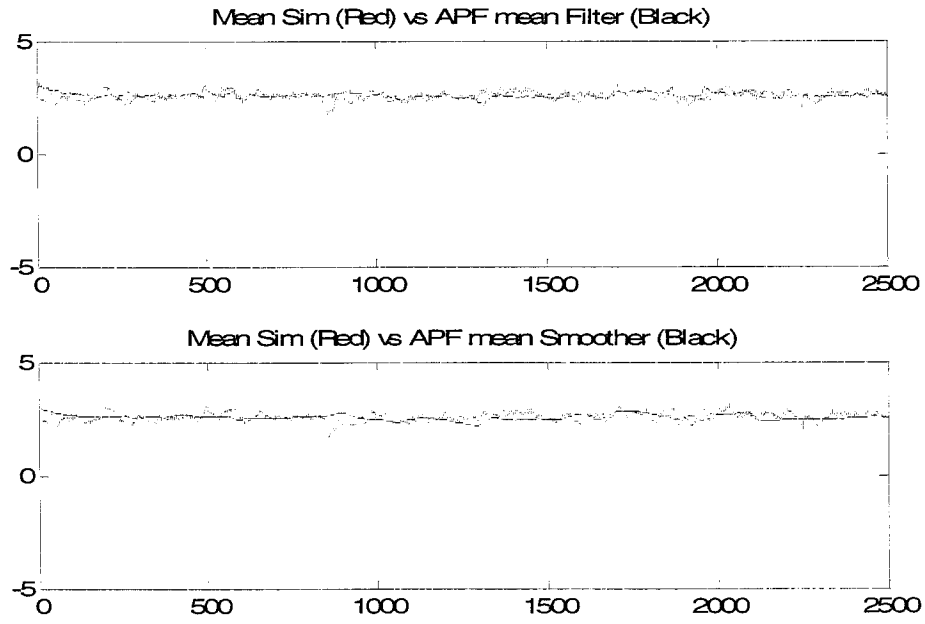


Figure 2.4b: SIR/APF Results on Simulations from SV3 Model

Along with the simulated mean (red lines), the corresponding estimation of the filtered and smoothed series are plotted. The results using the SIR algorithm and the APF algorithm are presented in panel (A) and panel (B), respectively.

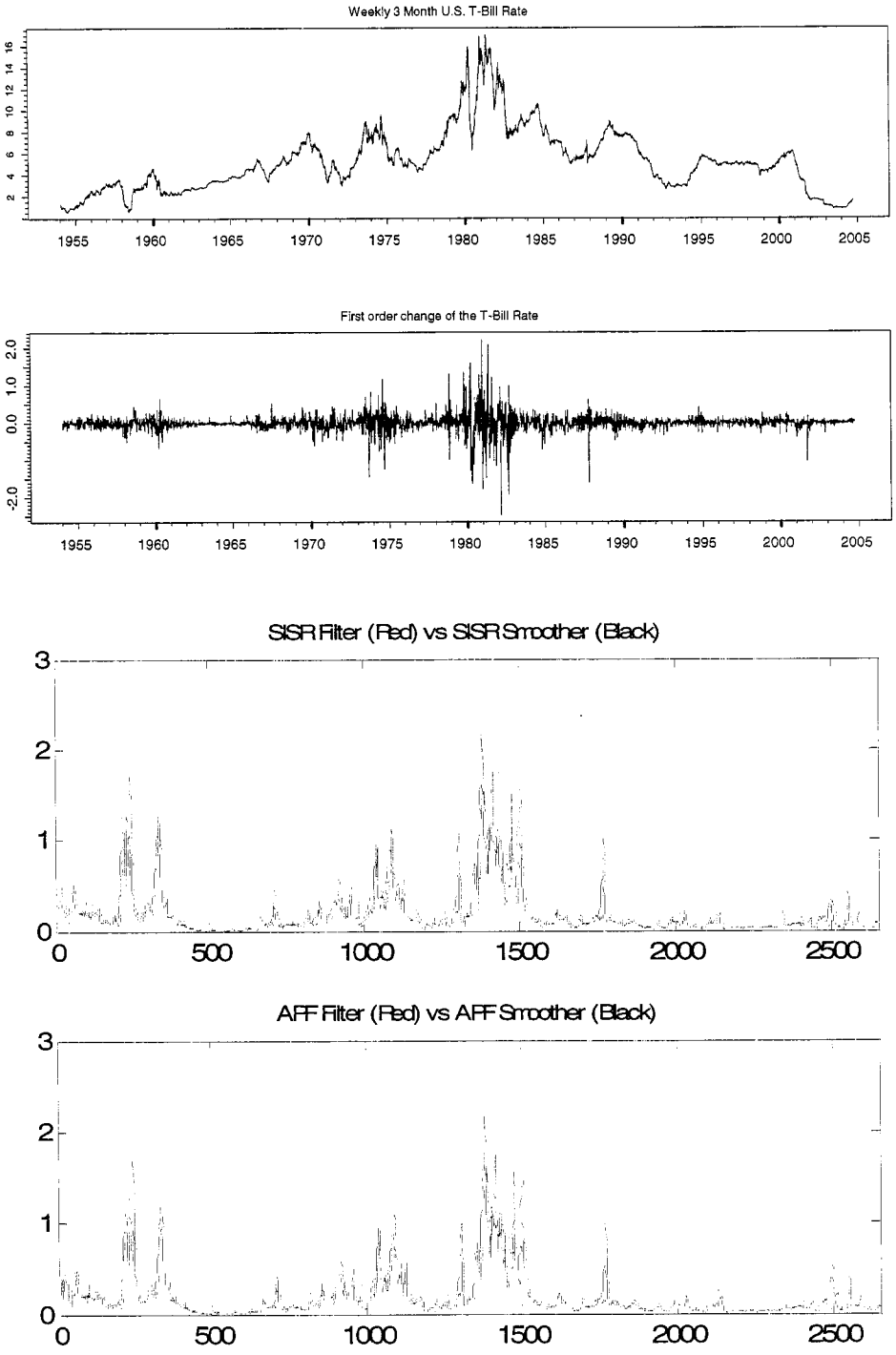


Figure 2.5: SIR/APF Results on US Short Rate: SV2 Model

2648 weekly observations of the 3-month T-Bill rates are plotted as well as its first-order difference. The filtered and the smoothed volatility process for the estimated SV2 model using the SIR and APF algorithm are presented accordingly.

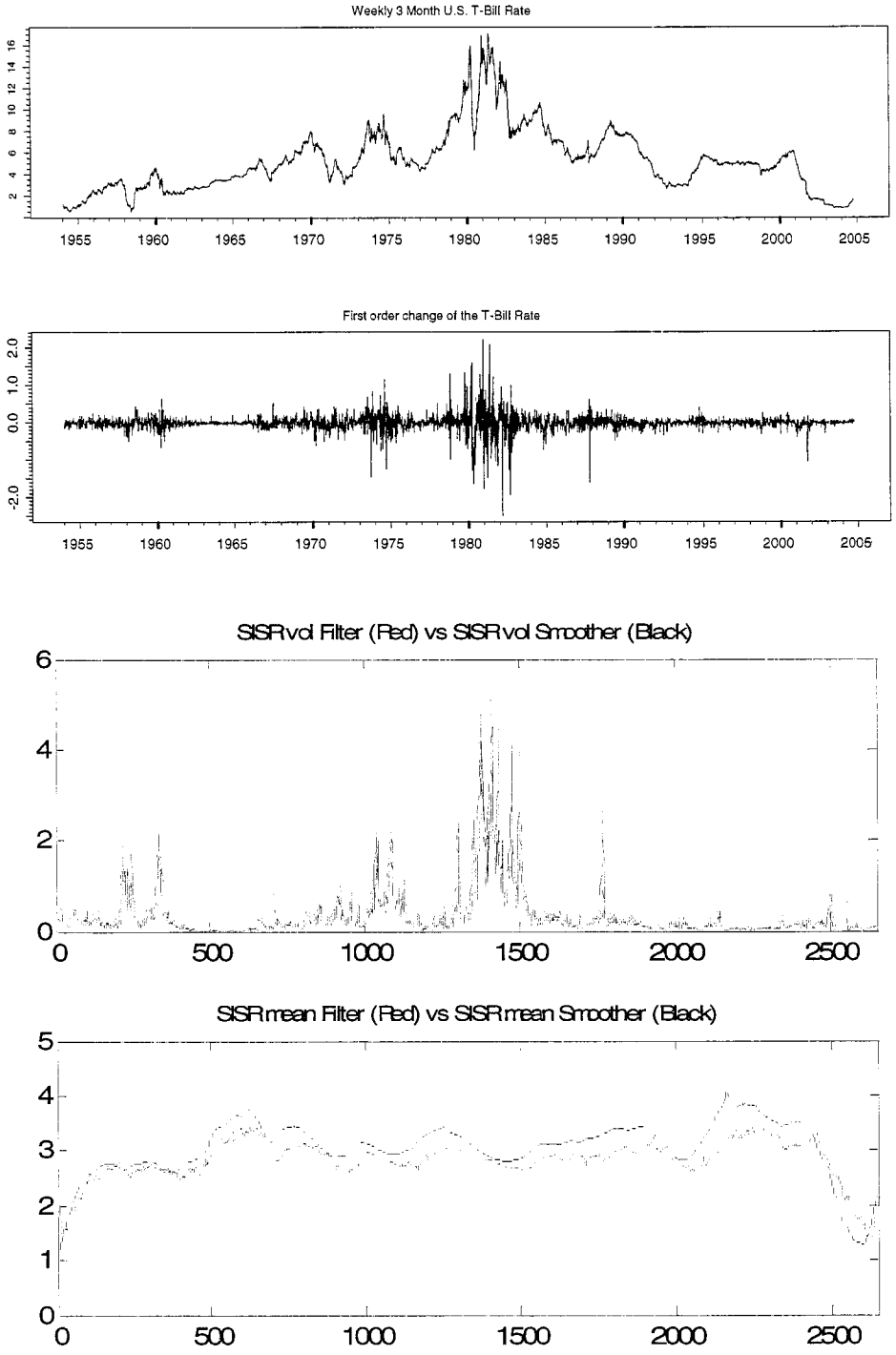


Figure 2.6a: SIR Results on US Short Rate: SV3 Model

2648 weekly observations of the 3-month T-Bill rates are plotted as well as its first-order difference. The filtered and smoothed mean/volatility processes for the SV3 model using the SIR algorithm are presented.

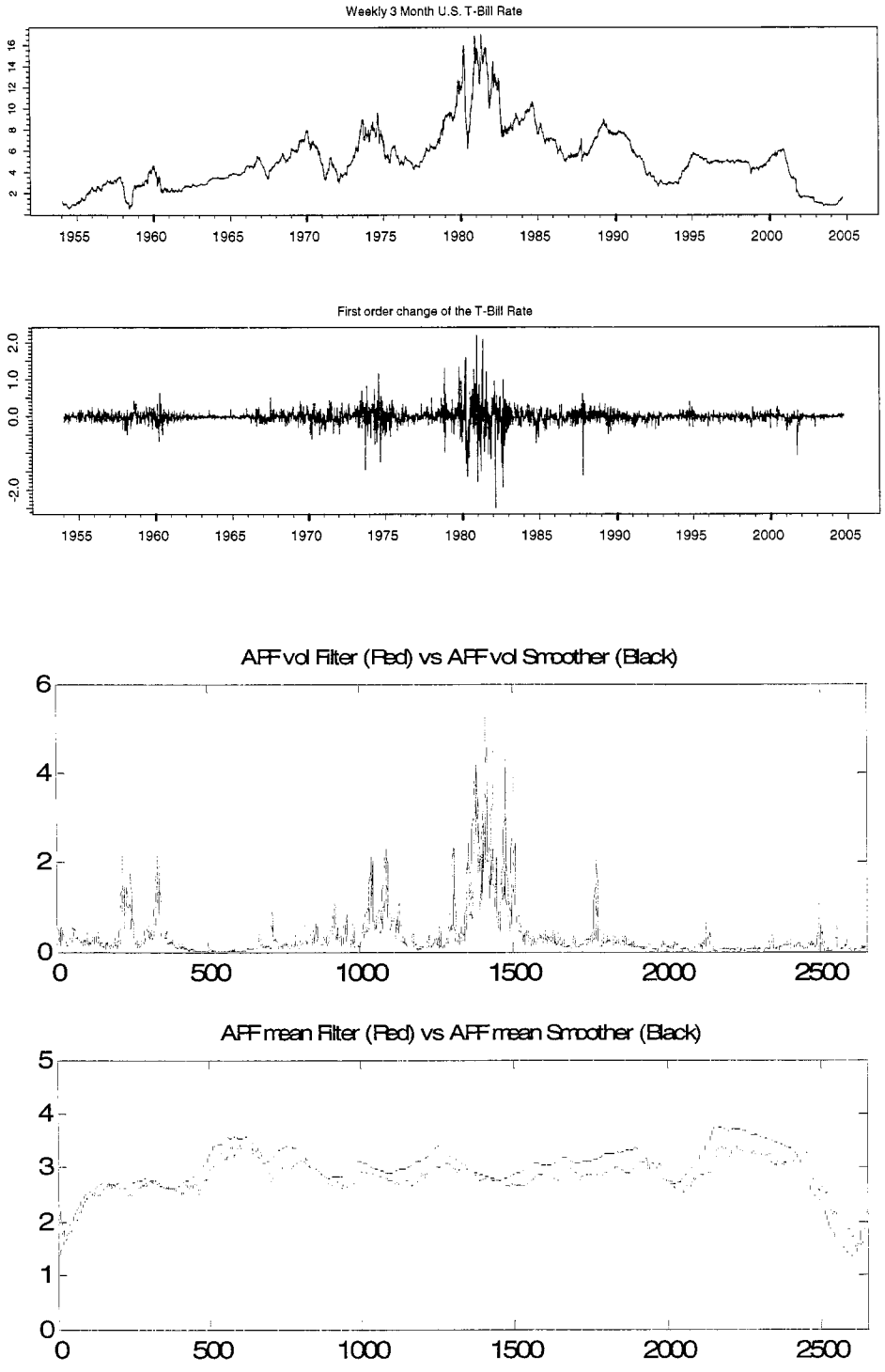


Figure 2.6b: APF Results on US Short Rate: SV3 Model

2648 weekly observations of the 3-month T-Bill rates are plotted as well as its first-order difference. The filtered and smoothed volatility processes for the SV3 model using the SIR and APF algorithm are presented accordingly.

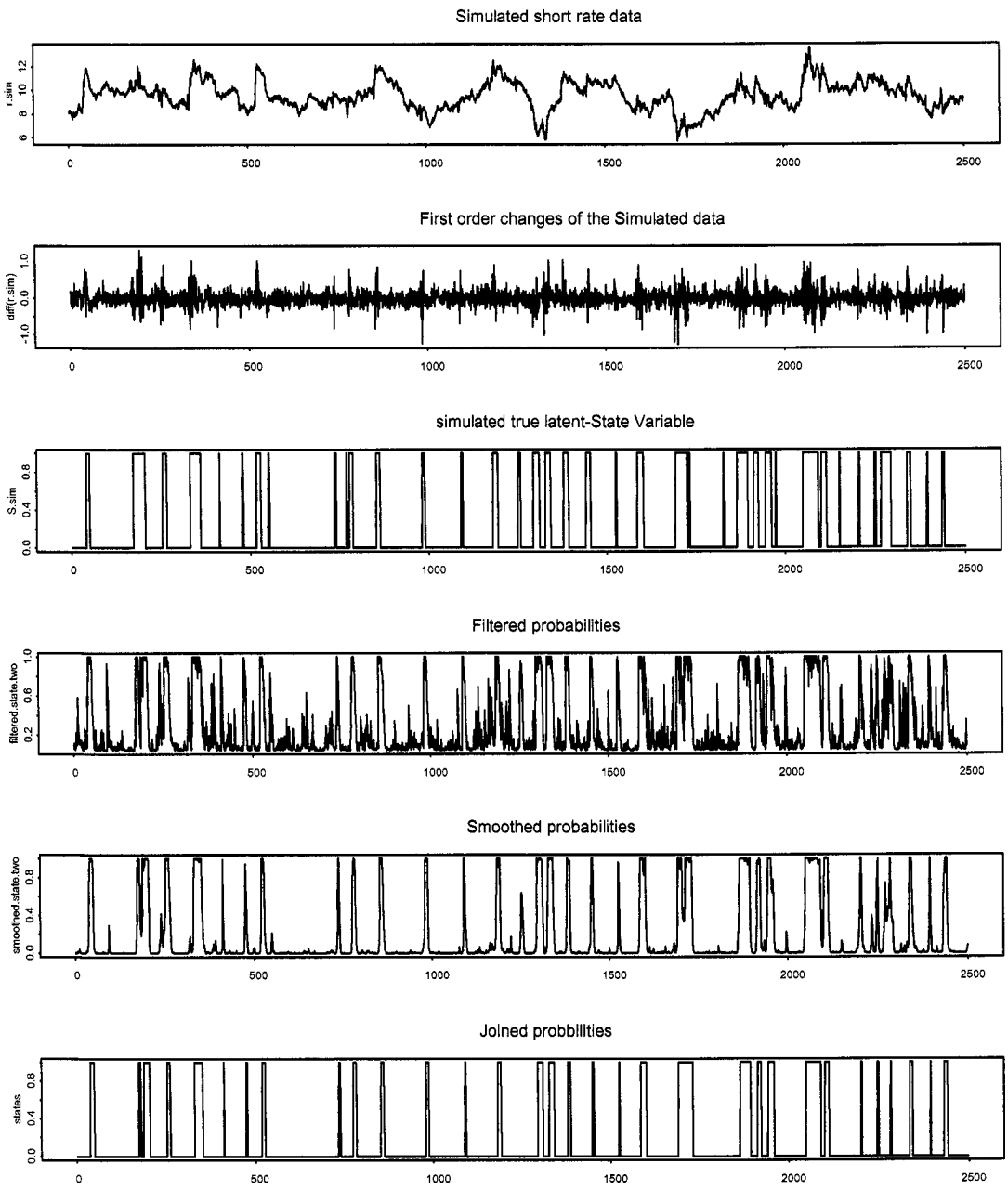


Figure 2.7: Results on Simulations from RS-in-σ Model

2500 observations are simulated from the RS-inσ model (2.4) with EMM estimates shown in Figure 1.9. The following panels plot (1) the simulated short rate series, (2) the first-order difference of the series, (3) the simulated state variable (0: low-volatility regime; 1: high-volatility regime), (4) the filtered probability and the smoothed probability for the state variable using the HMM algorithm, and (5) the sequence of the joint states using the Viterbi algorithm.

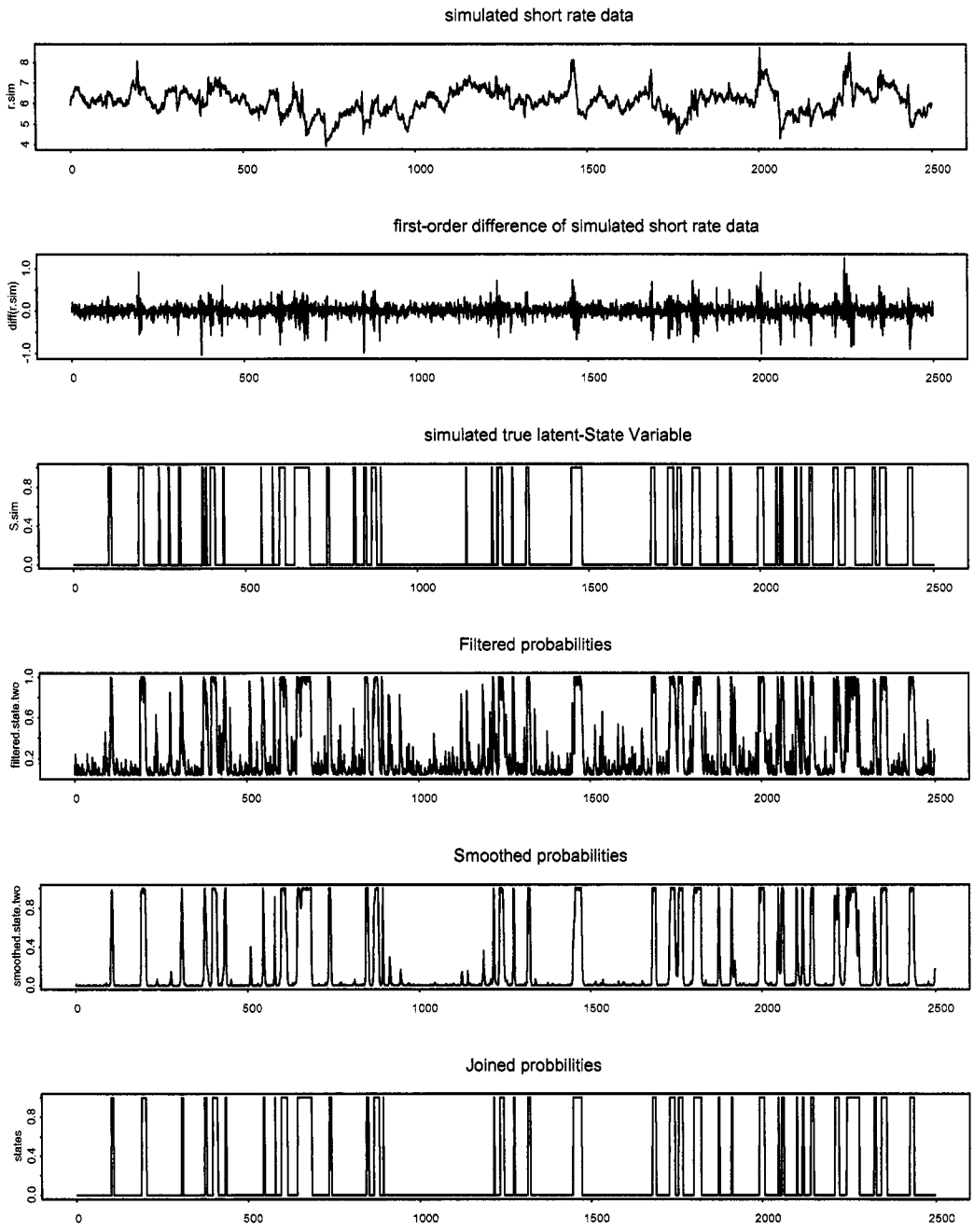


Figure 2.8: Results on Simulations from RS-in-$\sigma + Level$ Model

2500 observations are simulated from the RS-in-$\sigma + Level$ model (2.5) with EMM estimates shown in Figure 1.9. The following panels plot (1) the simulated short rate series, (2) the first-order difference of the series, (3) the simulated state variable (0: low-volatility regime; 1: high-volatility regime), (4) the filtered probability and the smoothed probability for the state variable using the HMM algorithm, and (5) the sequence of the joint states using the Viterbi algorithm.

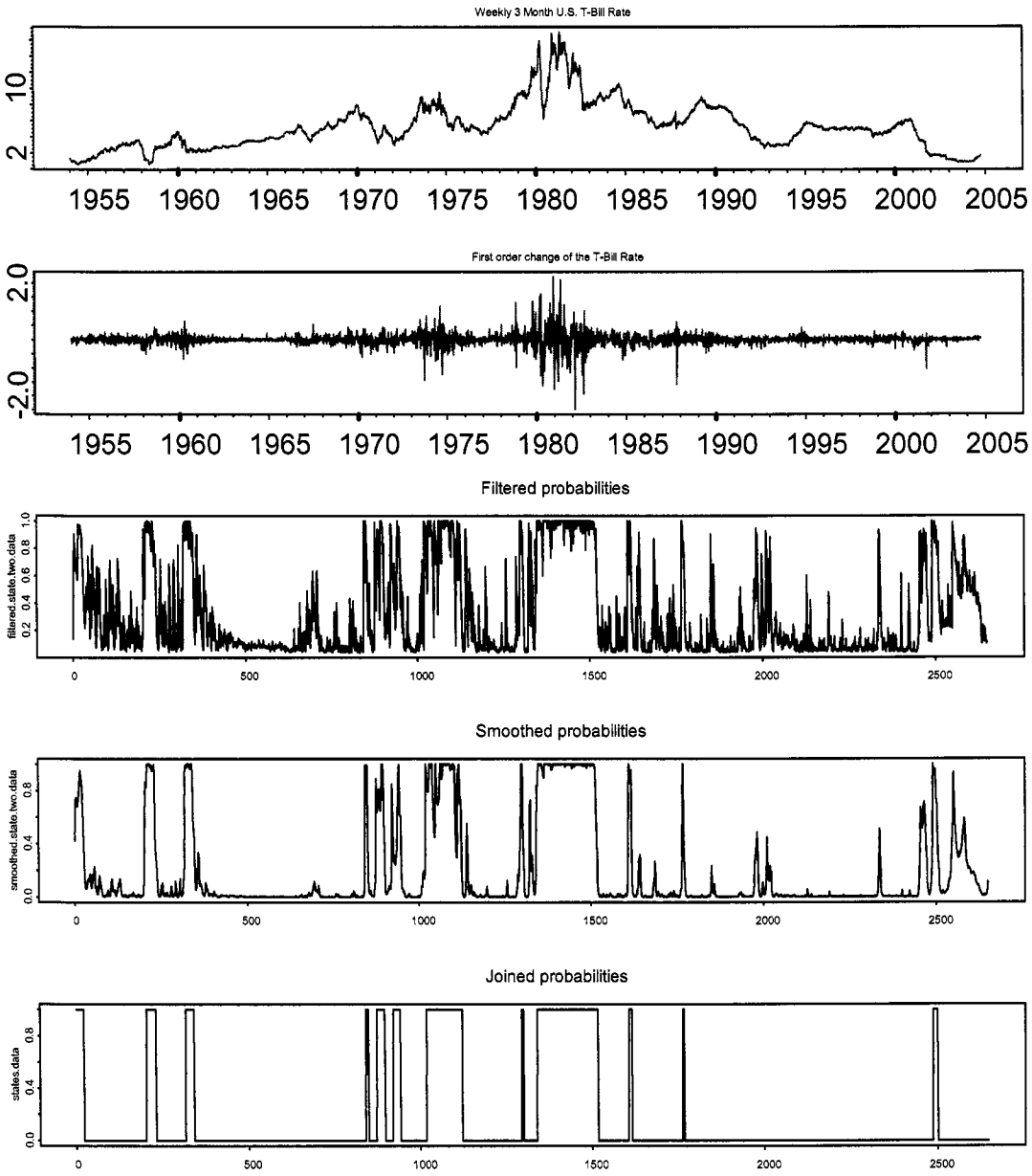


Figure 2.9: Results on US Short Rate: RS-in-σ Model

2648 weekly observations of the 3-month T-Bill rates, ranging from January 4, 1954 to September 24, 2004, are plotted as well as its first-order difference. The filtered and the smoothed probabilities, and the sequence of the joint states for the RS-in-σ model using the HMM algorithm are presented accordingly.

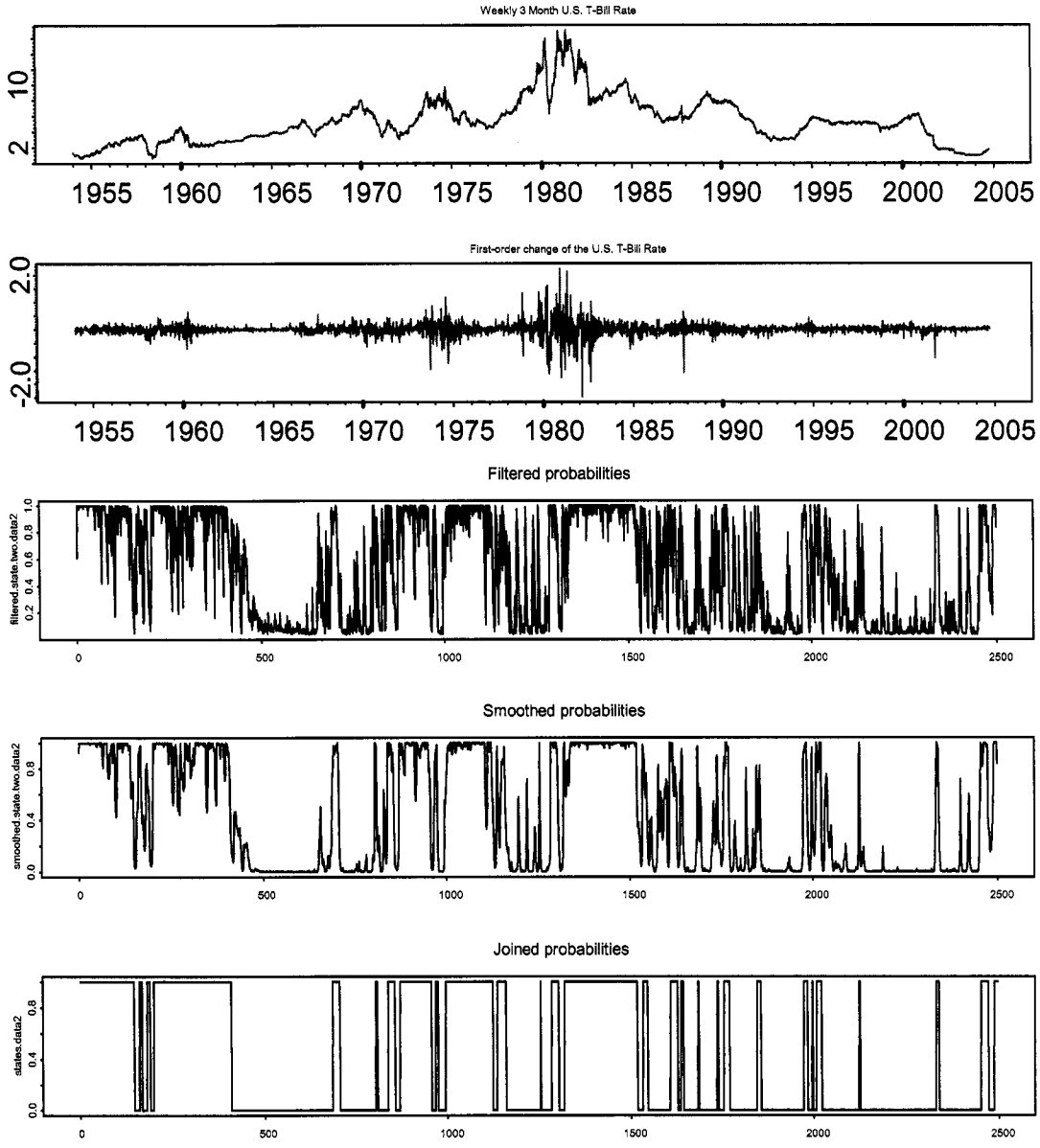


Figure 2.10: Results on US Short Rate: RS-in-σ + *Level* Model
2648 weekly observations of the 3-month T-Bill rates, ranging from January 4, 1954 to September 24, 2004, are plotted as well as its first-order difference. The filtered and the smoothed probabilities, and the sequence of the joint states for the RS-in-σ + *Level* model using the HMM algorithm are presented accordingly.

Table 2.1: Descriptive Statistics

The statistics summary is given for the raw data, 2648 weekly observations of the 3-month T-Bill rates, ranging from January 4, 1954 to September 24, 2004, and for the simulated data from the preferred candidate models.

Sample Moments	Mean	Std. Dev.	Skewness	Kurtosis
US short rate	5.246	2.849	1.065	4.712
SV2 model	1.987	1.657	1.418	5.629
SV3 model	2.636	1.572	2.204	14.407
RS-in-σ model	9.548	1.196	-0.098	3.157
RS-in-σ + *Level* model	6.091	0.620	-0.027	3.652

Table 2.2: Estimation Error

The root mean squared errors (RMSEs) for the estimated latent volatility process and/or mean process are reported for the SV models using the SIR algorithm and APF algorithm and for the RS models using the HMM algorithm.

T=2500	Latent volatility process		Latent mean process	
RMSE	**Filtered**	**Smoothed**	**Filtered**	**Smoothed**
SV2: SIR	0.6377	0.4976	-	-
SV2: APF	0.6368	0.4714	-	-
SV3: SIR	0.7194	0.5427	0.2134	0.2190
SV3: APF	0.7162	0.5467	0.2248	0.2319
RS-in-σ : HMM	0.2794	0.1775	-	-
RS-in-σ + *Level* : HMM	0.2491	0.1503	-	-

Chapter III: EMM Estimation of a Regime Switching Model: A Monte Carlo Study

3.1. Introduction

Models incorporating regime switching (RS) have a long tradition in time series econometrics. For example, early works using the switching regimes to model US real GDP growth include Hamilton (1989, 1990, and 1993), Filardo (1994) and Diebold and Rudebusch (1996). They showed that the Markov RS model provides a better fit to the data than a more complex, single-regime ARMA model. As pointed out by Hamilton (1989) and others, an RS model is more than a mere device used to fit the data; it has natural interpretations due to the association between the notion of regimes in the underlying econometric model and the observable economy-wide shocks that have persistent influence. In addition, RS models belong to a general class of mixture distributions. This feature enables them to generate a wide range of values of skewness, kurtosis and serial correlation even for two-state models. These properties inspire financial econometricians to apply RS models to financial time series. The RS behavior of financial series has been confirmed in the conditional mean dynamics of interest rates (Hamilton (1988), Gray (1996), Lak and Wang (1996), Sola and Driffill (1994), Ang and Bekaert (2002) and of exchange rates (Engle and Hamilton (1990)), and in the conditional variance dynamics of stock returns (Hamilton and Susmel (1994)) and of interest rates (Hamilton (1989), Garcia and Perron (1996), Gray (1996) and Ang and Bakeart (2001), Liechty and Roberts (2001)). The empirical evidence on interest rates have motivated the recent studies of the impact of regime shifts on term structure modeling (Naik and Lee (1997), Boudoukh et al. (1999), Evans (2001) and Bansal and

Zhou (2003), Dai, Singleton and Yang (2004)) and on bond pricing (Landén (2000) and Wu and Zeng (2003)).

Although there are voluminous empirical works using Markov RS models due to their theoretical and practical appeals, our knowledge on the efficient usage of them is still limited. A number of key issues have been hotly debated. Firstly, the questions of how to test the null hypothesis of no regime switches and how to identify the most appropriate number of regimes are not easily answered. Before modeling data with an RS model, one should verify the existence of multiple regimes and then test the optimal number of switching regimes. The difficulty in constructing such tests is that some of the parameters are not identified under the null hypotheses, which assumes single regime or fewer regimes scenarios. While many researchers have dedicated great efforts, mostly adopting the approach of Davies (1977, 1987), these problems are far from being settled; the tests that have been proposed are not very powerful and difficult to implement. Secondly, the test for the most appropriate RS model among many candidate specifications is rarely justified statistically. Due to the difficulty in answering the first issue, many (in fact most) authors assume the existence of multiple regimes and select the RS model in an ad hoc manner even when there is not theoretical guidance for which model specification is the appropriate one. For example, regime switches have been allowed in mean, and/or in volatility, and/or in autocorrelation coefficients in different empirical applications. Therefore, knowing how to compare different types of RS models becomes a very relevant issue. Lastly, it is often of interest to test between Markov and non-Markov RS models (an example is the threshold RS

model[35]), between discrete-time RS model and continuous-time model (for example, the multi-factor stochastic volatility (SV) model), and between hybrid RS models (for example, the mixture models of RS model and SV/GARCH model). Only limited estimation techniques are available to handle estimations and comparisons of these non-nested models and only a few papers dealt with these problems.[36]

As a consistent and efficient estimation technique, Efficient Method of Moments (EMM) is well known for its ability to estimate partially observable nonlinear systems, and for its special advantages in comparing non-nested model specifications.[37] To explore EMM's ability to estimate RS models may provide alternative insights for solving some of the inference problems discussed previously. However, not much is known about the inference of EMM for RS models; see only Bansal and Zhou (2002). This paper undertakes an extensive investigation of EMM estimation for an RS-in-volatility model through Monte Carlo techniques. This specific RS model has been carefully selected to accommodate the characteristics of financial series. We examine the efficiency sensitivity to the choice of the auxiliary models using ARCH and GARCH as the leading terms as well as nonparametric extensions; we compare the estimation efficiency to that of state space maximum likelihood (ML) procedure using Kim's approximate filter (1994); we investigate the appropriateness of the over-identification test for goodness-of-fit and analyze the estimation results when the

[35] See McAleer (1995) for an overview.
[36] For example, Pesaran and Potter (1992) used a sophisticated simulation technique, extended by M. Pesaran and B. Pesaran (1993), to test between Markov and threshold models.
[37] See for example, Andersen and Lund (1997) for interest rate modeling, Dai and Singleton (2000) for term-structure modeling, Chernov and Ghysels (2000) for option valuation, etc..

structural models are not correctly specified. The finite-sample properties of EMM efficiency in small samples are also examined.

The contribution of such a Monte Carlo study is two-folded. Firstly, the investigation of whether EMM estimation can readily accommodate general extensions to the functional form and distributional assumption of regime switches will provide additional evidence for the effectiveness of EMM estimation as a generally applicable methodology for RS models.[38] Secondly, the conclusions in this paper are likely to be applied to other types of RS models characterized by regime shifts not just in the volatility dynamics. Our results will help provide guidelines for obtaining desirable EMM estimates for RS models through appropriate estimation designs. In addition, the reliability of EMM inference in the RS contexts will spur the research in this direction and provide new insights to many issues that have been hotly debated recently.

The remainder of the paper is organized as follows. A brief description of the EMM methodology is provided in Section 2. Section 3 analyzes the properties of several commonly RS models. Section 4 illustrates the selected Markov RS-in-Sigma model and describes the Monte Carlo design. Section 5 reports the empirical results and section 6 summarizes and concludes.

3.2. EMM Estimation

The theory of EMM estimation is developed in Gallant and Tauchen (1996) and is extended to non-Markovian data with latent variables in Gallant and Long (1997). The basic procedure of EMM estimation consists of two steps. First, the empirical

[38] See Andersen al et. (1999) for a Monte Carlo study of a SV model using EMM.

conditional density of an observed time series is estimated by a seminonparametric (SNP) series expansion. This SNP expansion has an AR-(G)ARCH Gaussian density as its leading term, and departures from the Gaussian leading term are captured by a Hermite polynomial expansion. This step is accomplished by projecting the data onto a SNP model and is therefore called the projection step. Second, a GMM-type criterion function is constructed using the score functions (from the log-likelihood of the SNP density) as moments. The scores are evaluated using the simulation output from a given structural model and the criterion function is minimized with respect to the parameters underlying the structural model. This step is to extract structural parameters from the data summary by minimizing the χ^2 criterion and is therefore called the estimation step. The expository discussion of the method is in Gallant and Tauchen (2001). A summarized detail of the ideas is presented in Chapter 1.

3.3. The Regime Switching Model

The RS model has been utilized to better accommodate the time-varying conditional volatility and to capture the skewness and fat tails of financial time series. Because of its popularity, more and more attention has been paid to the characteristics of different RS models. Timmermann (2000) provided the closed-form calculation of higher order moments for several types of RS models. In this section, we summarize the findings in Timmermann (2000) especially those on the third and fourth moments of different types of RS models. Based on these stylized facts, a RS model is selected for our Monte Carlo experiment and characteristics of the simulated series from this RS model are provided.

Without loss of the generality, Timmermann (2000) considered three types of Markov RS models:

RS1: $$y_t = \mu_{S_t} + \sigma_{S_t}\varepsilon_t \tag{3.1}$$

RS2: $$y_t = \mu_{S_t} + \beta y_{t-1} + \sigma_{S_t}\varepsilon_t \tag{3.2}$$

RS3: $$y_t = \mu_{S_t} + \beta_{S_t} y_{t-1} + \sigma_{S_t}\varepsilon_t \tag{3.3}$$

where $t = 1,..,T$, ε_t is i.i.d. N(0,1), and the state indicator, $S_t = 1,..,k$, is assumed to be governed by a first-order Markov process with the transition probabilities from regime i to regime j at time t is defined as $p_{ij} = \Pr(S_t = j \mid S_{t-1},...,S_1) = \Pr(S_t = j \mid S_{t-1} = i)$ with restriction of $\sum_{j=1}^{k} p_{ij} = 1$ for all i's.

Each of the above models has been widely adopted in applied econometric studies. For example, the RS1 model, allowing regime switches in mean and variance, has been used in empirical work by Engel and Hamilton (1990) to analyze growth in US GDP. State-independent and state-dependent autoregressive dynamics allowed in the RS2 and RS3 have been used by Gray (1996) for fitting the US short-term interest rate. Stylized facts on these three types of RS processes in Timmermann (2000) are summarized as follows:

- A shift in the mean can produce very substantial persistence around the unconditional mean, even when there is no autocorrelation dynamics for y_t. The autoregressive effect is an increasing function of the size of the shift.[39] In addition, a simple model with shifts only in the mean appears not sufficient to generate

[39] According to Perron (1990), the autocorrelation coefficient with rare shift gets closer to one.

excess kurtosis.

- A shift in the variance, on the other hand, can produce strong persistence around the unconditional variance, and generate relative large excess kurtosis. However, a necessary condition for the Markov RS process to generate skewness is the state-dependent means; state-dependence in the variance alone is insufficient.
- When shifts in the mean and variance are both allowed, the generated skewness, excess kurtosis, and autocorrelations are not just simply mixed up; very different patterns may be displayed depending on the size of the shifts.
- Once an autoregressive dynamics of y_t is introduced as for the RS2 and RS3 model, both the skewness and kurtosis become larger for specific pairs of the transition probability. In this setting, even low-order autoregressive Markov RS model with a small number of states can provide a basis for very flexible econometric models.
- The transition probabilities are very important in determining the higher-order moments and the patterns of the serial correlation for the RS process. On one hand, the scope of asymmetry and fat tails that can be generated is extremely expanded; on the other hand, it becomes much harder to understand the complexity when applying RS models.

While many RS models are readily handled by EMM estimation, we focus on the following autoregressive RS-in-sigma model for our Monte Carlo experiment:

$$y_t = \mu + \beta y_{t-1} + \sigma_{S_t} \varepsilon_t \tag{3.4}$$

where $(\mu, \beta, \sigma_1, \sigma_2, P_1, P_2) = (1, 0.9, 2, 4, 0.95, 0.90)$ is assumed as the parameter specification for a two-state case, and the evolution of the state indicator S_t is given by

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} = \begin{pmatrix} P_1 & 1-P_1 \\ 1-P_2 & P_2 \end{pmatrix} \tag{3.5}$$

Well used in many empirical works, this model is characterized by: (1) regime-independent mean and autoregressive dynamics: the change of the series is assumed to mean revert to level 10 with the reversion speed $(1-\beta = 0.1)$; and (2) regime-dependent volatility dynamics: there is 95% probability for the series to stay in the low-volatility regime ($\sigma_1 = 2$) and 90% probability in the high-volatility regime ($\sigma_2 = 6$).The above parameter specification is selected to accommodate the characteristics of an interest rate series, such as the US short rates. For example, the mean-reverting coefficient is very close to[40], but still less than, unity, ensuring that y_t is stationary and ergodic, which is consistent with many empirical finding. Also, the transition probabilities are very persistent with both exceeding 0.90, which is in line with the past RS evidence for financial series dynamics. In addition, the constraint on conditional volatilities and transition probabilities of $(\sigma_1 < \sigma_2, P_1 > P_2)$ is imposed to keep consistent with the literature, which implies that the low-volatility regime is more persistent.

1000 representative observations are simulated from this discrete-time RS model described by equation (3.4) and (3.5) using the function *simulation.MSAR* in S-PLUS. Although this RS model is relatively simple, it is good enough to generate complex

[40] Recent regime switching research points out this may be due to the model misspecification of single-regime models.

behavior as that implied from the financial series. Table 3.1 shows the descriptive statistics for the simulated series and its first-order different series; Figure 3.1 presents the simulation plot and the ACF plots for the level and the squared series. Basic characteristics of the simulated data from this RS model are: nonstationary behavior (slow mean reversion), large changes and small changes are clustered together (ARCH effect), and excess kurtosis (non Gaussian distribution). These characteristics that are created by the regime-depended variance and autoregressive process in the RS model (3.4) are consistent with the stylized facts about the characteristics of the financial series. Notice that the simulated series report fat tails but modest skewness, which is consistent with what has been pointed by Trimmermann (2000) that the excess kurtosis is necessarily generated by the regime-dependent variance, but the skewness is necessarily by the regime-dependent mean.

3.4. Monte Carlo Design

Given the above specification, the Monte Carlo experiment has been performed as follows. Firstly, the *observed* data of a specific sample size (T=500, T=1000, T=4000) is simulated using the *true* parameter set. Secondly, the simulated series is projected onto an auxiliary model with the (G)ARCH model as a leading term. For better selections, we utilize a set of starting values to explore the global maximum of the quasi-likelihood function, which is suggested for higher-order SNP models by Gallant and Tauchen (2001). Note that we did not perform complete searches of the BIC-preferred SNP model for each single simulation due to the concerns of computing time. Instead, we completely searched the best SNP model for only one simulation, and

assumed the corresponding functional form as the preferred SNP model for all simulations.[41] Therefore, the parsimonious structure of SNP model is pre-selected; the parameters of SNP model are re-estimated for each simulation. Thirdly, a long series of length N=50,000 is generated from the *true* structural model (3.7) with the *true* parameter set as the starting values. The simulated series is then used to construct the score function. Fourthly, the underling parameters for model (3.7) are estimated with a two-step procedure: (1) a grid search is adopted first over a range of values for the transition probabilities with the other parameters fixed at their true values. A pair of transition probabilities is chosen, for which the EMM objective value is the lowest. (2) The EMM objective function is then minimized with respect to the other parameters with the transition probabilities fixed at the selected pair. Lastly, information regarding other aspects of EMM inference in estimating the RS model is collected. For each combination of score generator and sample size, 500 Monte Carlo experiments are performed; all the simulation and estimation are conducted in S-PLUS using the FinMetrics 2.0 module.

Several issues related to the Monte Carlo design and the EMM estimation design are underscored as follows. Firstly, we adopt a grid searching for estimating the transition probability because the EMM objective is not a continuous function of the transition probability, which results in interrupted iterations with respect to these parameters. Figure 3.2 provides the sample surface of the EMM objective statistics obtained from a grid search for the pair of the transition probability. Clearly shown by the graph, the EMM objective function is discontinuous when a small change in the

[41] Actually, this simplification is feasible; the results of completely searching of the best SNP model for many simulations have pointed the same functional form of the SNP model.

transition probability occurs. For this case, if we estimate the transition probability together with the other parameters using the usual EMM procedure, the estimates of the transition probability will barely change from its starting values. Although the estimates for the other parameters can be obtained, they may end up at inefficient estimation points when the EMM iteration is interrupted. To deal with this discontinuity issue, a grid search is easy to implement and has been used by many EMM applications in RS contexts; see Chernov, et al (2003) and Bansal and Zhou (2003) for examples.[42] Secondly, the two-step grid-searching procedure stated above may not be feasible in practice due to the fact that we usually do not know the true parameters. A more practical procedure is to minimize the EMM objective function with respect to all the other parameters for each fixed pair of the transition probability and choose the pair, and the associated estimates for the other parameters, that give the overall lowest EMM objective value. The reason that we still use the procedure stated above is because that it can lower the computing load significantly without losing much estimation efficiency relative to that of using the more practical procedure. The optimal grid searching procedure is of interests for future research. Thirdly, the SNP model is argued to be capable of approximating virtually any smooth distribution, even a mixture distribution, which is the case for a model with regime shifts. However, unlike some model specifications, the specification of the SNP model for RS models is not clear; that is, one may not have a priori score generator, which is expected to be preferable.[43] The moments analyses based on the steady-state transition probabilities in Trimmermann

[42] Andersen, Benzoni, and Lund (2002) smoothed the jumps to make the objective function continuous.
[43] For example, Nelson (1990) showed that the EGARCH models and the stochastic volatility models share the identical continuous-record diffusion limit.

(2000) appear to indicate that an ARCH/GARCH type of SNP models with nonparametric parameters may be appropriate for our RS model; for example, Bansal and Zhou (2002) used a VAR(1)-ARCH(5)-Kz(4)-Kx(2) SNP model for their RS models in the term structural study. But for the EMM estimation, a larger SNP model is needed for a longer series, so we still consider to select the ARCH-leading and GARCH-leading SNP models based on the BIC criterion. Other types of auxiliary models are left for future research. Lastly, according to guidelines provided by Zivot and Wang (2005), sometimes better convergence can be obtained by transforming the parameters so that the parameters to be estimated are unconstrained. Therefore, we estimate the transformed parameters (β, P_1, P_2) to make sure they are between zero and unity.

3.5. Empirical Results

The performance of EMM estimation using the preferred (G)ARCH-leading SNP models is reported in Section 3.5.1 for different sample sizes. Similar information for the state space ML estimates using Kim's filter (1994) is discussed in Section 3.5.2 as convenient comparisons. Complementary evidence regarding EMM inference is also provided along various dimensions, including the quality of its goodness-of-fit specification test in Section 3.5.3, and the capability of identifying misspecified structural models in Section 3.5.4

3.5.1. Choice of the Auxiliary Model:

Selecting a good auxiliary model is important for EMM estimation. According to Gallant and Tauchen (2001), only as the auxiliary model approaches the true conditional density, will the estimated covariance matrix for structural parameters approach that of ML. Consequently, the asymptotic efficiency of the EMM estimator will approach that of ML. Gallant and Tauchen (2001) recommended the SNP model as the auxiliary model and as the score generator. They argued that the SNP model is capable of approximating virtually any smooth distribution, even a mixture distribution, which is the case for a model with regime shifts. Therefore, in this section, we present a sensitivity study of EMM estimation based on the choice of the SNP model. Specifically, we consider an ARCH-leading and GARCH-leading SNP model selected by the BIC criterion. Other types of auxiliary models are left for future research.

For the limited sample size T=500, we firstly select the preferred SNP model; more parsimonious score generators are strongly favored by the BIC criterion. Specifically, the best SNP model with an ARCH-leading term is an AR(1)-ARCH(2) model (10210000) and the GARCH-leading one is an AR(1)-GARCH(1,1) model (11110000). Allowing non-parametric parameters such as (Kz and Kx) is not beneficial for both SNP models, where we have experienced either prohibitively long convergence times or non-convergences due to numerically unstable QML estimations. This is consistent with the finding in Andersen (1999) for EMM estimation of SV models. After the projection step, we establish the grid for searching the transition probabilities. A coarse search is performed with the other parameters fixed at their true values; most of the estimates fall into the area of (0.8, 1) for P_1 and of (0.5, 1) for P_2. To lower the

computing load, we consider an equal-space 10-by-10 grid for the searching; that is, the estimation nodes in the grid are at (0.81, 0.83, …, 0.99) for P_1 and at (0.5, 0.55, …, 0.95) for P_2. Using the selected pair of transition probabilities after searching the above grid, we then estimate the other parameters using EMM estimation.

Table 3.2 displays the results using two alternative score generators (10210000 and 11110000), including the mean, 5% quantile, 95% quantile, and associated root mean squared errors (RMSEs) of the EMM estimates; Figures 3.6 plots the corresponding distribution for each parameter. Although slightly lower RMSEs are obtained for estimating (μ, β, P_1, P_2) using the GARCH-leading SNP model, the two score generators perform similarly for estimating these parameters. For both score generators, the transition probability of staying in the high-volatility regime, P_2, is estimated less accurately than that of staying in the low-volatility regime, P_1. Specifically, the estimation of P_2 has a long left tail away from its true value of 0.9. On the other hand, the ARCH-leading SNP model presents much better estimation for (σ_1, σ_2) than the GARCH-leading SNP model. Specifically, using the AR(1)-ARCH(2) SNP model, the mean estimations of (σ_1, σ_2) are less biased, their associated RMSEs are lower, and the distributions are more centered. The distribution plots of $(\mu, \beta, \sigma_1, \sigma_2)$ and histogram plots of (P_1, P_2) in Figure 3.3 give a clear display of the above discussion. This finding implies that an ARCH-leading SNP model may better capture the volatility process involving regime shifts than a GARCH-leading SNP model when the sample size is small, although many EMM empirical works have shown uniformly the advantage of using the GARCH-leading SNP model for other

types of model specifications.

It is intuitive that more moments from the SNP model are required to capture the characteristics of larger data samples. For T=1000, including a nonparametric fat-tail parameter (Kz) is now beneficial for the GARCH-leading SNP model: an AR(1)-GARCH(1,1)-Kz(4) model (11114000) is selected based on the BIC criterion; but a relative parsimonious ARCH-leading SNP model, AR(1)-ARCH(3) model (10310000), is still preferred. Using the same searching grid for the transition probabilities, the results for the two score generators above are shown in Table 3.3. When the sample size expands, the mean estimates are closer to the true values and the associated root mean squared errors (RMSEs) of all estimates become uniformly smaller than when T=500, implying improvement of EMM efficiency for larger sample sizes. For example, the transition probabilities are estimated more precisely for this sample size, although the estimation of P_2 is still less accurate than that of P_1 for both score generators. In terms of the relative performance of two score generators, they perform very similarly for estimating (μ, β); the RMSEs are very close. The RMSEs obtained for estimates of (P_1, P_2) are close too, but their histogram plots in Figure 3.6 show that the GARCH-leading SNP model seems to generate more centered estimates. As for estimates of (σ_1, σ_2), the ARCH-leading SNP model, again, provides lower RMSEs, especially for the high-volatility estimate of σ_2. However, the estimate for the low-volatility, σ_2, is slightly less biased using the GARCH-leading SNP model. The corresponding distribution plots in Figure 3.4 nicely illustrate the results.

For the relative large sample size T=4000, the dimensions of the SNP model

expand significantly; a higher degree of nonparametric fat-tail parameter (*Kz*) is now needed for both SNP models. An AR(1)-GARCH(1,1)-*Kz*(6) model (11116000) and an AR(1)-ARCH(5)- *Kz*(6) model (10516000) are preferred by the BIC criterion. Using the same searching grid for the transition probabilities, the performance of EMM estimation greatly improves using either score generator. As shown in the Table 3.4, the mean estimates are extremely close to the true values and the associated RMSEs are the smallest, implying that EMM is quite reliable for large samples. In particular, the estimates for the transition probabilities are clearly improved. Moreover, the results for estimates (μ, β) appear roughly independent of the leading term in the SNP model. The GARCH-leading SNP model begins to show advantages over the ARCH-leading SNP model for estimating the volatility-related parameters in this sample size. The RMSEs of estimates (P_1, P_2) obtained from the GARCH-leading SNP model are lower and the corresponding distributions appear more centered, as illustrated in the histogram plots in Figure 3.5. In addition, the GARCH-leading SNP model provides more accurate estimates for the volatility parameters (σ_1, σ_2), although it was inferior to the ARCH-leading SNP model when estimating these RS volatility parameters in smaller samples.

3.5.2. Comparison of EMM Relative to ML

An alternative method to estimate dynamic linear Gaussian state space models involving unobserved variables is maximum likelihood (ML) based on the Kalman filter. A variety of extended filters has been proposed for non-linear and non-Gaussian systems such as the RS model; see Hamilton (1988, 1989), Cosslett and Lee (1985), and Harrison and Stevens (1976). Among them, Kim (1994) presented his filtering and

smoothing algorithms for a Markov RS state space model[44], along with ML estimation of the unknown parameters. As an extended version of the Kalman filter and the Hamilton filter with appropriate approximations, Kim's filter allows for inferences in a large class of Markov RS models as long as they can be put in the state space form. Because of this advantage, Kim's filter has been widely applied in many RS applications and therefore serves as a good benchmark to evaluate the EMM estimation results. The brief idea of the Kim's filter is outlined in the Appendix B.

The means, quantiles, and RMSEs of the state space ML estimates using Kim's filter are provided in the bottom panels of Tables 3.2-3.4 for T=500, T=1000, and T=4000, respectively; Figure 3.6 shows the improvement of ML estimators when sample size expands. For the sample size T=500 or T=1000, EMM is not as efficient as the ML-based method. Although the RMSEs of the EMM estimates for the volatility parameters (σ_1, σ_2) are smaller than those of ML estimates, the ML produces much less biased estimates for (σ_1, σ_2). In addition, the EMM encounters much difficulty in estimating the transition probability of staying in the high-volatility regime (P_2). The efficiency of EMM estimation and ML estimation improves substantially for T=4000; the associated RMSEs become extremely small for this longer, but still empirically relevant, sample especially for financial data. Although the RMSEs of the EMM estimates are still larger than those of ML estimates, the improvements verify the reliability of EMM inference and suggest that the efficiency of EMM approaches that of the ML-based procedure in this setting. In addition, EMM is a general applicable

[44] State space models are useful tools for expressing dynamic systems that involve unobserved state variables; see Harvey (1989) and Hamilton (1994a, 1994b) for expositions.

method that can be used for more models than ML, while it is much slower to estimate the RS model, and many other types of model specifications, than Kim's approximation filter.

3.5.3. Specification Test

In the estimation step of EMM, parameters of a candidate structural model are obtained by minimizing the EMM objective function using the fitted auxiliary model as the score generator. If the structural model is correctly specified, then the product of the sample size and optimized EMM objective function is χ^2 distributed with degree of freedom equal to the difference between dimensions of the auxiliary model and the structural model. This statistic therefore provides a GMM-type over-identification test for the model adequacy.

In our Monte Carlo experiments, we can investigate the performance of the EMM over-identification test in the RS contexts. If the finite-sample distribution of the EMM test statistic conforms to its asymptotic distribution, the p-values should be asymptotically uniform over the fractiles. Therefore, the plot of the fractions of p-values that fall within each fractile will help show how well the finite-sample EMM test statistics conform to their asymptotic distribution. In particular, the 0-5% and 5-10% fractiles provide the possible biases of the goodness-of-fit tests at the (asymptotic) 5% and 10% level. Similar analyses have been provided in Andersen et al. (1996) and (1999) for estimating the SV model using GMM and EMM, respectively.

Figure 3.7 displays the fraction plots for p-values that fall within the indicated 5% fractile using the corresponding preferred SNP models for different sample sizes. The

findings on the over-identification tests of EMM are not quite satisfactory. For a given leading term of the SNP model, the fraction plots suggest that the size distortion of the EMM statistics is growing significantly as the sample size increases. Specifically, the frequency observed in the 0-5% fractile is 4% for T=500 when the GARCH-leading SNP model is utilized. However, the EMM statistics will reject at the 5% level about one quarter of time for T=4000. At the same time, the mass located in the right-tail fractiles tends to decrease as the sample size expands. That is, the tendency of EMM test statistics is to over-reject the true structural model when the sample size is relative small, but as the sample size expands, the statistics are badly inflated. The same phenomenon is observed for both score generators. The reason for the increased size distortions is a mystery. However, Andersen and et al. (1996) reported similar results for the specification test when estimating the SV model using GMM, in which the choice of the moment function is subjective and thus inefficient for some cases. Our results here about the size distortion of the EMM statistic could be a signal that the estimated covariance matrices obtained from the BIC-preferred auxiliary models is imprecise and other types of SNP models should be explored in the future research.

Another finding is that, for each sample size, the pattern of the p-value fraciles is quite similar, although not exactly the same, across the two score generators. Therefore, there is no advantages of one over the other as the leading term of the SNP model. The preferred ARCH-leading SNP model for T=1000 actually produces the most appropriate specification test. However, the result is misleading because the better performance may be just due to the fewer moments used in this preferred ARCH-leading SNP model for this specific sample size. If this is the case, one may have to pay more attention in

practice to the tradeoff between information and precision involved in the choice of the auxiliary model for the EMM procedure in the RS contexts. That is, for similarly performing SNP models, the usage of a score generator requiring a fewer number of moments may be beneficial for the accuracy of the specification test of EMM.

3.5.4. Choice of the Structural Model:

The performance of EMM estimation is presented in previous sections when the *true* structural model, an RS-in-sigma model, is correctly specified in the second step of EMM. However, in most of cases, we don't know the true process generating the observed series, so we often compare different types of model specifications in terms of in-sample fitting and out-of-sample forecast. For example, the Markov RS model has recently been advocated as a successful alternative to the SV model for describing financial series due to its interpretation flexibility and its capability of generating a wide range of values for skewness, kurtosis and serial correlation. In Chapter 1, we showed the success of an RS model as well as a two-factor and a three-factor SV model for fitting the weekly US short-term interest rate; they all could not be rejected by EMM objective function at 10% significant level. However, the three accepted models have different implied behaviors. Therefore, a reasonable question is whether EMM is capable of identifying and rejecting the misspecified structural models and accepting correct ones.

In order to investigate this issue, we estimate by the EMM a misspecified structural model although the *true* data is simulated from model (3.4). The misspecified model we consider is a two-factor continuous-time SV model given by:

$$\begin{cases} dy_t = (\mu - (1-\beta) y_t)dt + \sigma_t dW_1 \\ d\log(\sigma_t^2) = (\omega_0 + \omega_1 \log(\sigma_t^2))dt + \xi dW_2 \end{cases} \tag{3.6}$$

where dW_1 and dW_2 are mutually independent Wiener processes. The SV2 model is an extended and continuous-time version of the model (3.4) with the introduction of a latent variable. Its key characteristic is that the change of y_t mean-reverts towards the long-run level $\mu/(1-\beta)$, with the reversion speed of $(1-\beta)$.[45] In addition, the log-volatility of series is assumed to follow a mean reverting process. The shock to the log-volatility is measured by the non-negative parameters ξ.

Table 3.5 reports the mean, the 5% quantile, and the 95% quantile of the EMM estimates for the misspecified two-factor SV model by (3.6) based on two BIC-preferred SNP models (11114000 and 10310000) for the true data with sample size (T=1000). For each auxiliary model, the p-values based on the χ^2 distribution associated with the EMM test statistic are also saved. There are over 325 out of 500 Monte Carlo experiments, for which, the p-values lead to rejections of the two-factor SV model for both ARCH and GARCH leading SNP models based on significant level 5%. Specifically, the rejection rate is over 65% for either SNP model. This finding verifies that EMM is capable of capturing the failure of the two-factor SV model when fitting the volatility clustering that is actually created by an RS model. The fact that EMM can distinguish a misspecified structural model is encouraging, while one may need to run additional specification tests when evaluating the model misspecification in different settings. In addition, it is worthy to point out the difference between the

[45] The conditional variance of Δy_t does not depend on y_t for simplification.

estimates from these two model specifications. The process from the RS model (3.4) is a mean reverting process to level 10 with speed of adjustment 0.1, and the variances switch between 2 and 4. While a strong mean reverting process with similar long-run reverting mean is also implied from the accepted SV2 model, the magnitudes of the estimates for the mean dynamics from the SV2 model are significantly different from the true parameters from the RS model. The means and the ranges for each parameter are similarly estimated using the preferred auxiliary models with ARCH and GARCH as the leading term.

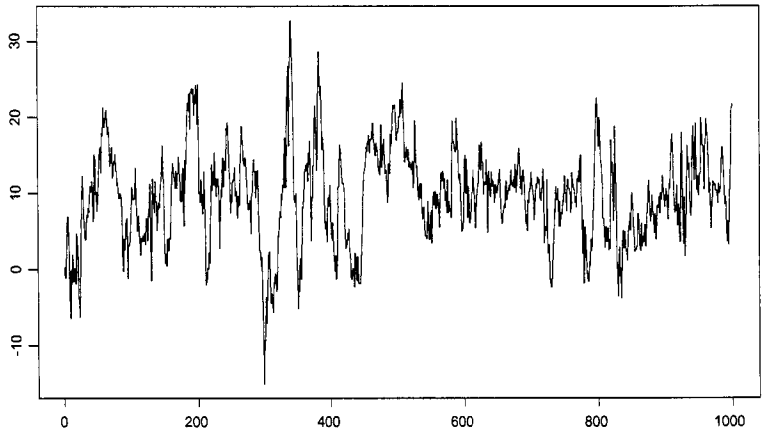
3.6. Conclusion

In this paper, we perform an extensive Monte Carlo study on a selected RS-in-volatility model to evaluate EMM estimation. Although EMM is generally not as efficient in small samples as the state space ML procedure using Kim's filter (1994), its efficiency improves greatly and approaches that of the likelihood-based estimation when the sample size expands. In addition, the EMM approach is flexible to estimate many types of RS models, even those that can not be put into the state space form, which is a required condition for using Kim's filter in the ML procedure.

Several aspects of EMM estimation of the RS model are investigated, and we provide the following observations and suggestions regarding its implementation. Firstly, the estimation efficiency in different sample sizes is examined based on the choice of the SNP model with either an ARCH or a GARCH as the leading term. We find that the dimension of the BIC-preferred SNP model increases as the sample size expands; parsimonious parameterization of the SNP model is preferred for smaller

samples. Although the estimation efficiency becomes relatively robust to the choice of the leading term, the ARCH-leading SNP model appears to be more reliable for estimating the regime-switching volatility parameter in smaller samples. Many empirical works that utilize EMM estimation considered the GARCH model as a more efficient leading term, but our results imply that the GARCH-leading SNP model shows advantages over the ARCH-leading model only for larger samples. Secondly, our Monte Carlo setting facilitates an investigation of the quality of the finite-sample over-identification test of EMM estimation in the RS context. We notice that this goodness-of-fit test, which performs remarkably well in the SV context as shown in Andersen et al. (1999), is less reliable for inference in the RS context. The size distortion based on the fraction of the p-value fractiles increases with the sample size. One may have to make keen decisions on the tradeoff between information and precision when choosing the appropriate SNP model as the moment generators. Thirdly, the capability of EMM to identify misspecified structural models is briefly analyzed. The rejection rate based on 5% significant level is over 65% for the misspecified two-factor SV model, implying that EMM is capable of capturing its failure for fitting the volatility clustering that is actually created by an RS model.

(A)



(B)

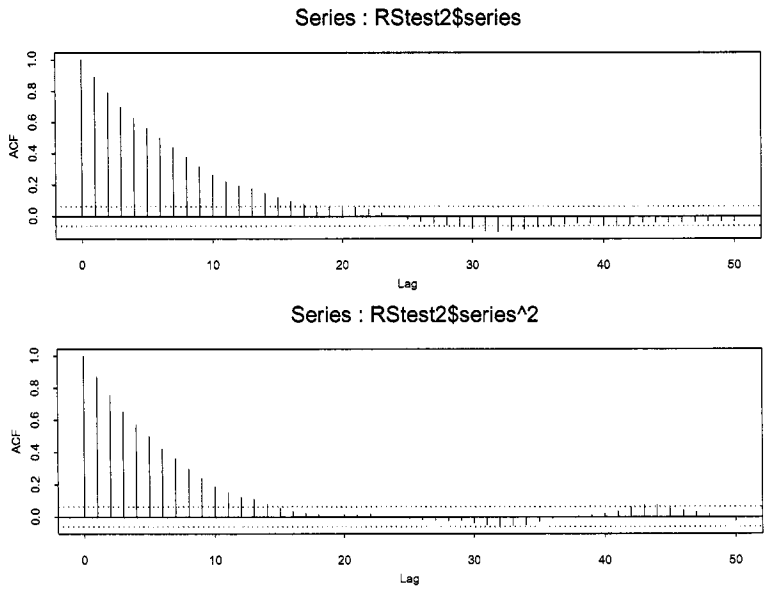
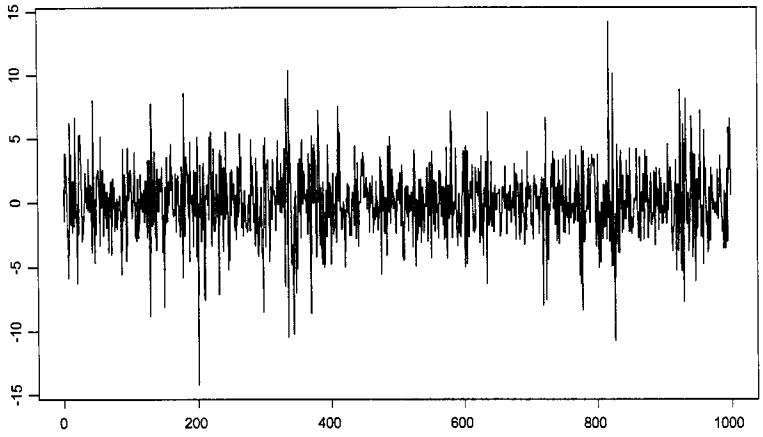


Figure 3.1a: Simulated Series from the RS Model

1000 observations are simulated from the RS model (3.4) with parameter set $(\mu, \beta, \sigma_1, \sigma_2, P_1, P_2) = (1, 0.9, 2, 4, 0.95, 0.9)$. The series and the ACF for the level and the squared of the series are presented in panel (A) and (B).

(C)



(D)

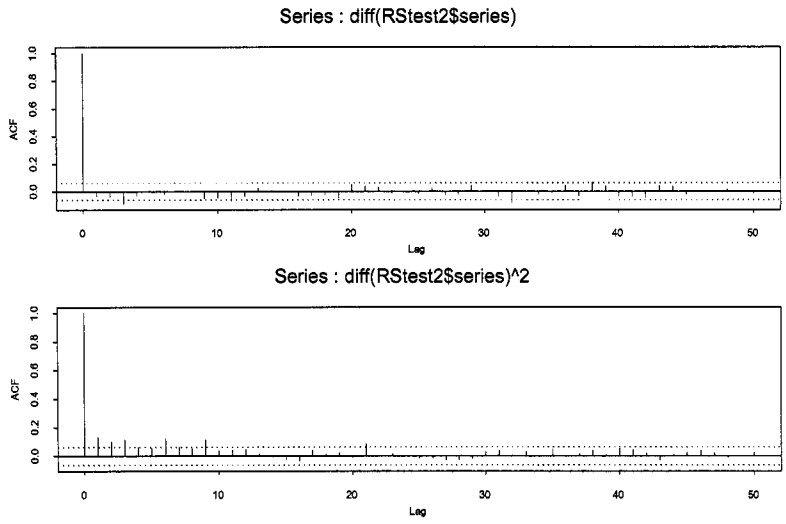


Figure 3.1b: Difference of Simulated Series from RS Model

1000 observations are simulated from RS-in-sigma model (3.2) with parameter set of $(\mu, \sigma_0, \sigma_1, P_1, P_2) = (1, 2, 4, 0.9, 0.9)$. The first-order difference of the series and the associated ACF plots are presented in panel (C) and (D).

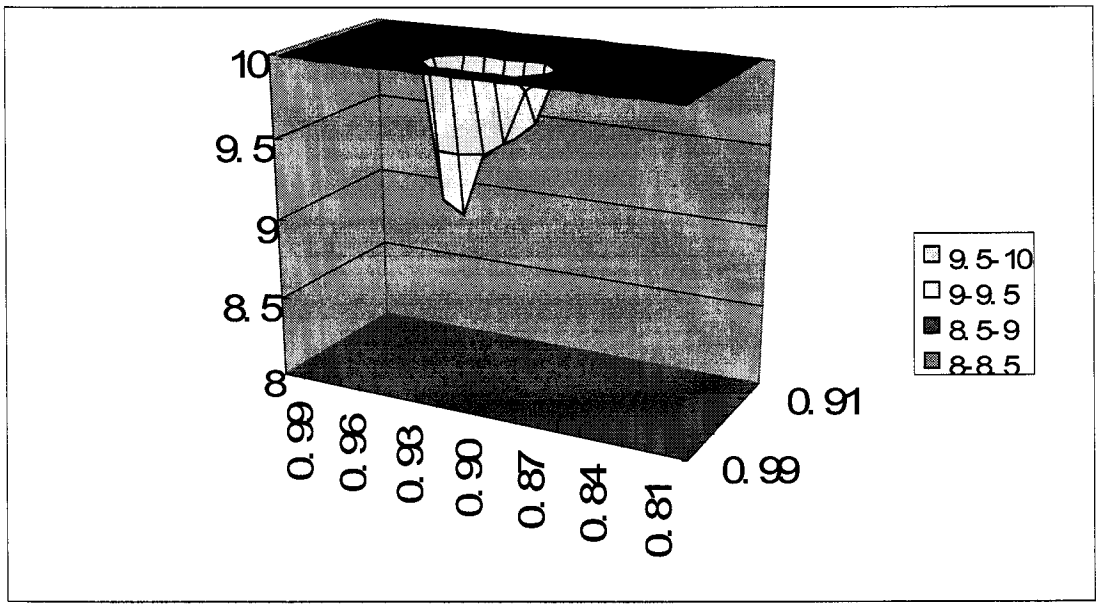


Figure 3.2: EMM Objective Statistic Surface

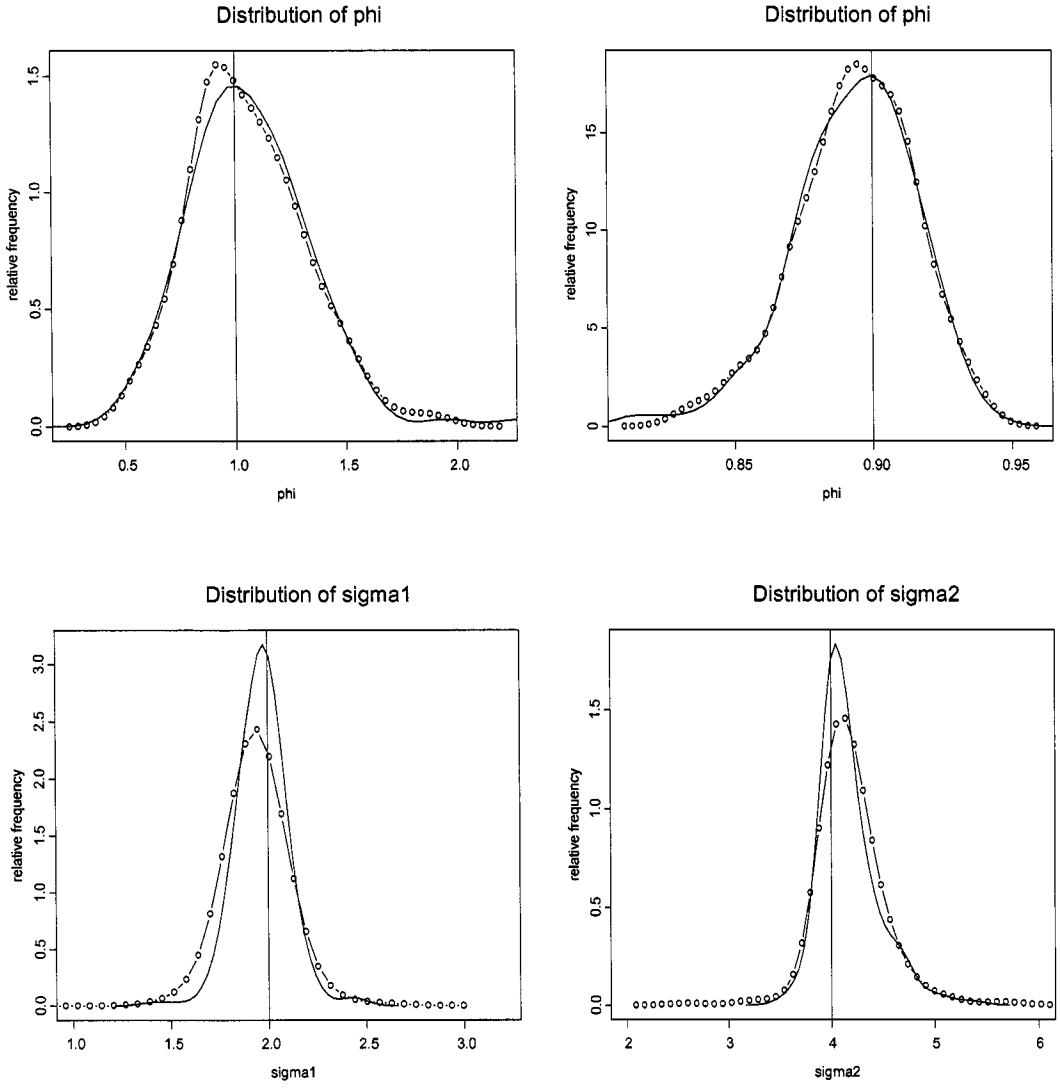
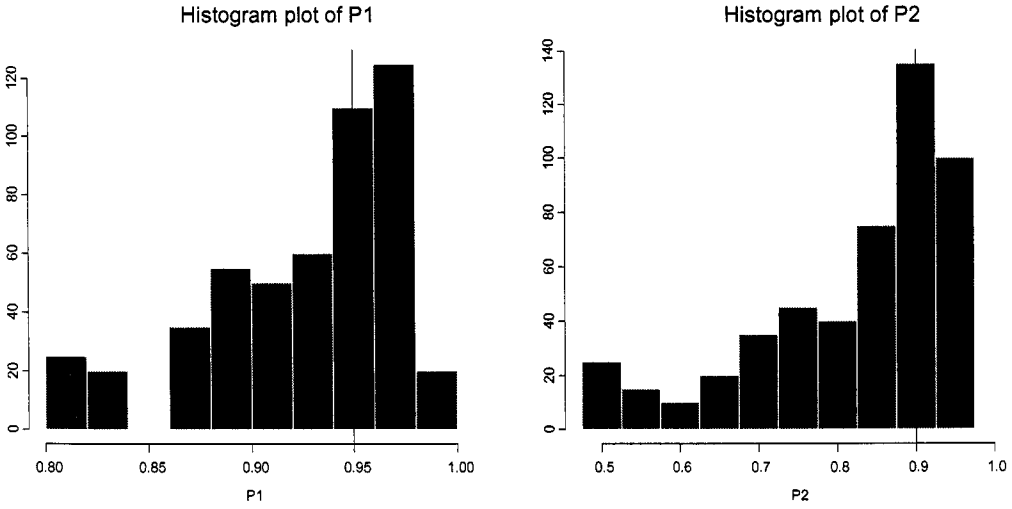


Figure 3.3a: Distribution of EMM Estimates (T=500) I

The distributions of EMM estimates $(\mu, \beta, \sigma_1, \sigma_2)$ are plotted using the BIC-preferred ARCH-leading SNP model 10210000 (solid lines) and GARCH-leading model 11110000 (dotted lines) with vertical lines as the associated true values.

(A) ARCH-leading SNP



(B) GARCH-leading SNP

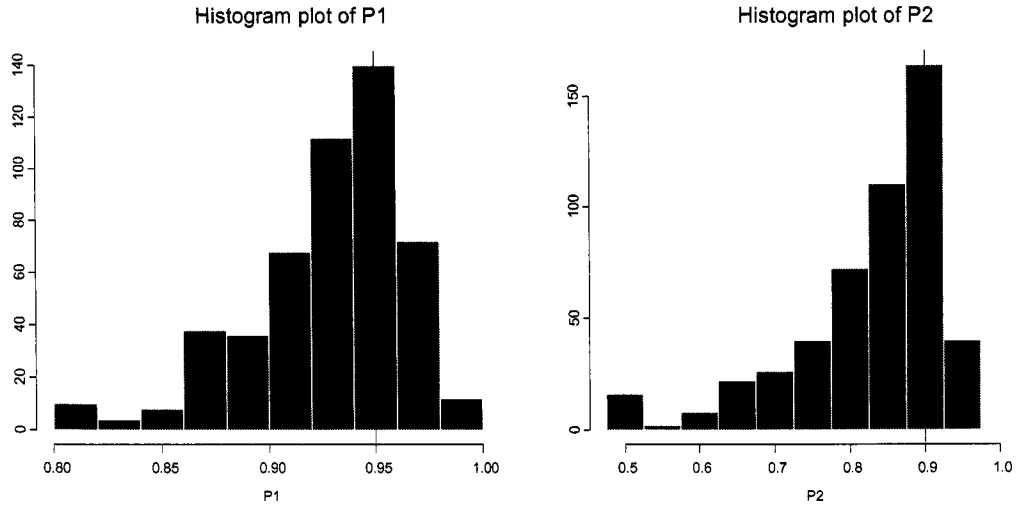


Figure 3.3b: Distribution of EMM Estimates (T=500) II

The histograms of EMM estimates (P_1, P_2) are plotted using the BIC-preferred ARCH-leading SNP model 10210000 (Panel A) and GARCH-leading model 11110000 (Panel B) with vertical lines as the associated true values.

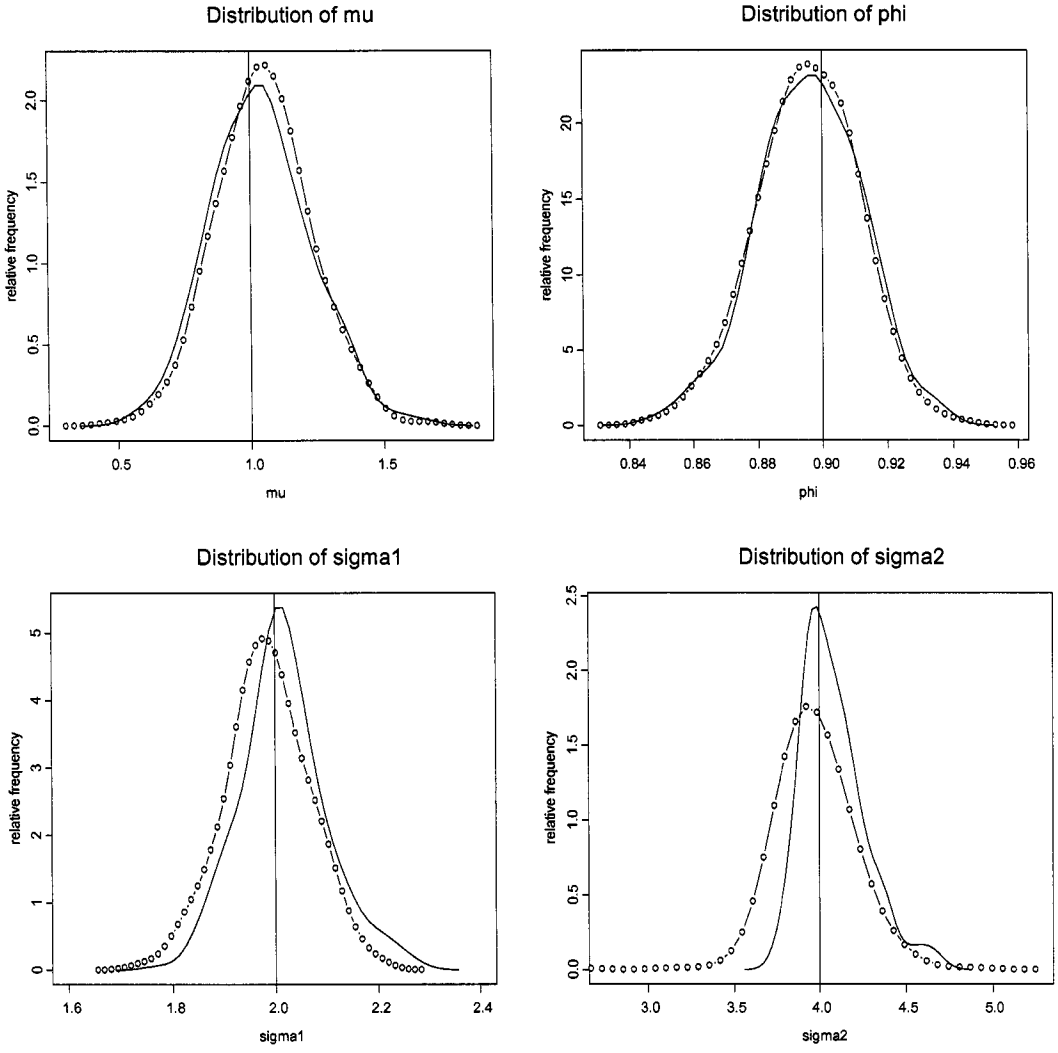
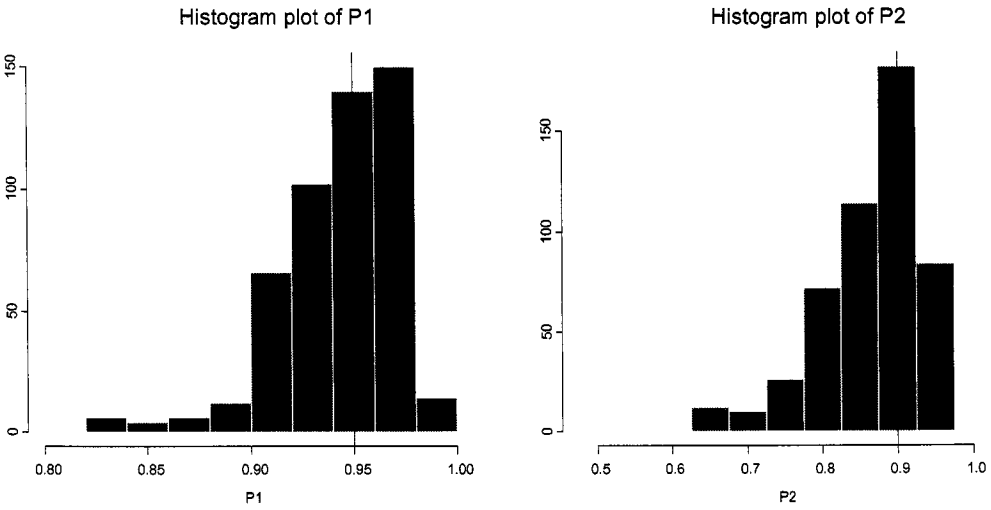


Figure 3.4a: Distribution of EMM Estimates (T=1000) I

The distributions of EMM estimates $(\mu, \beta, \sigma_1, \sigma_2)$ are plotted using the BIC-preferred ARCH-leading SNP model 10310000 (solid line) and GARCH-leading model 11114000 (dotted line) with vertical lines as the associated true values.

(A) ARCH-leading SNP



(B) GARCH-leading SNP

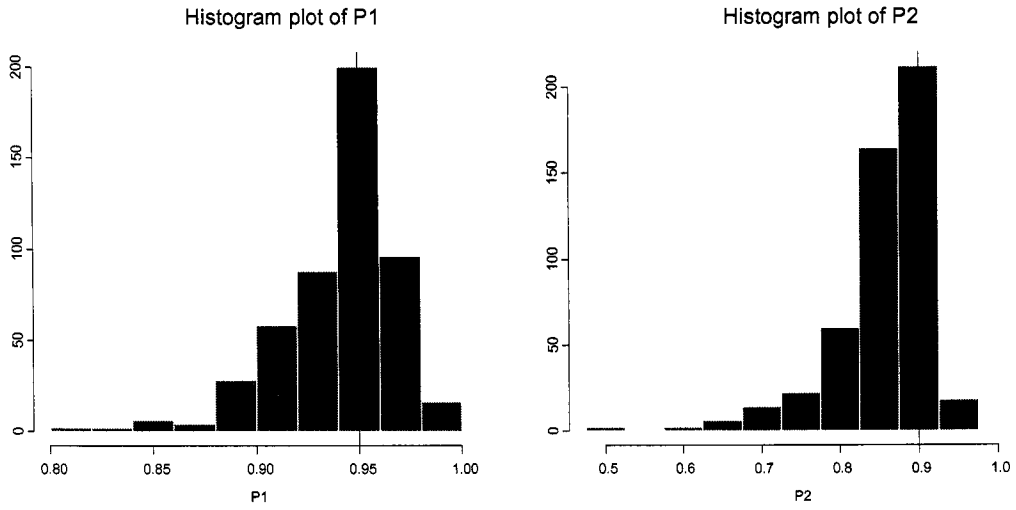


Figure 3.3b: Distribution of EMM Estimates (T=1000) II

The histograms of EMM estimates (P_1, P_2) are plotted using the BIC-preferred ARCH-leading SNP model 10310000 (Panel A) and GARCH-leading model 11114000 (Panel B) with vertical lines as the associated true values.

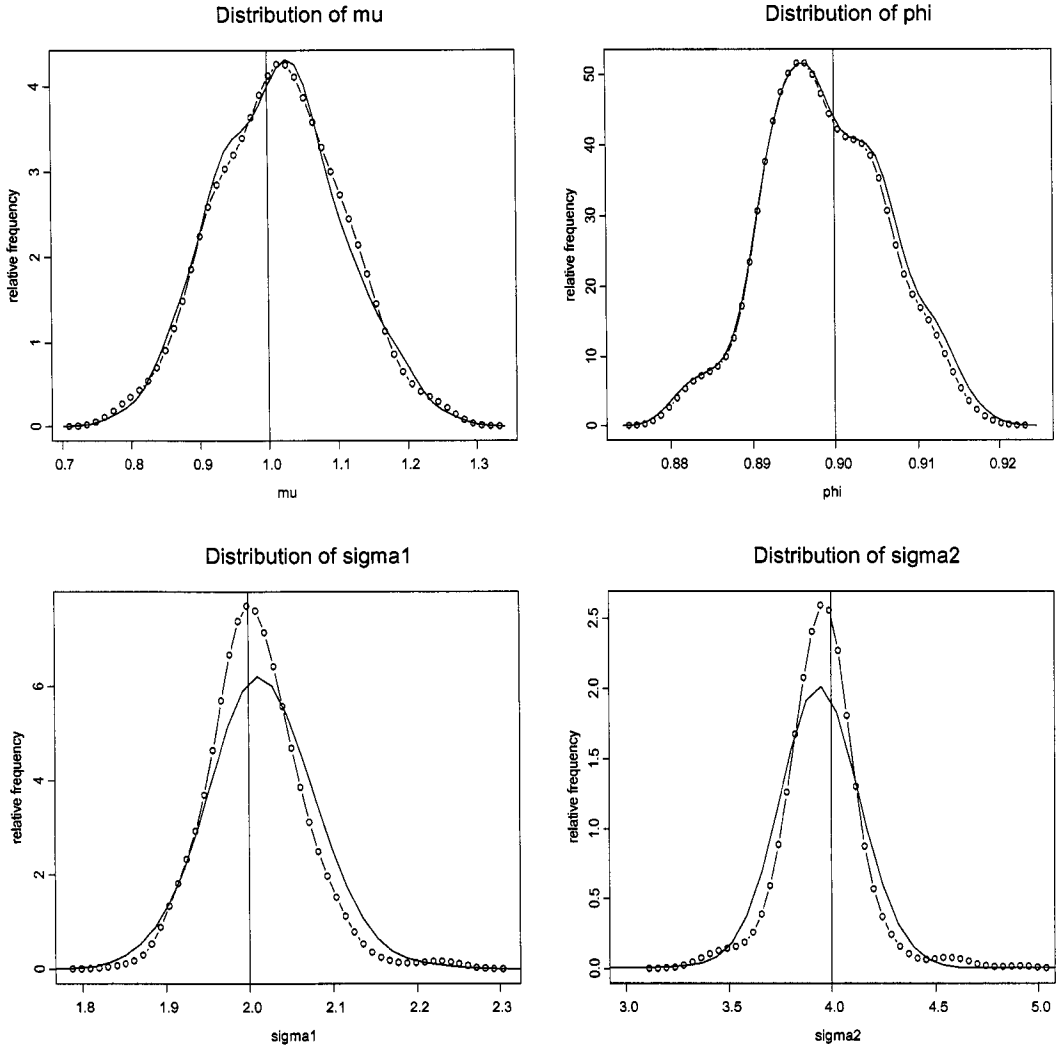
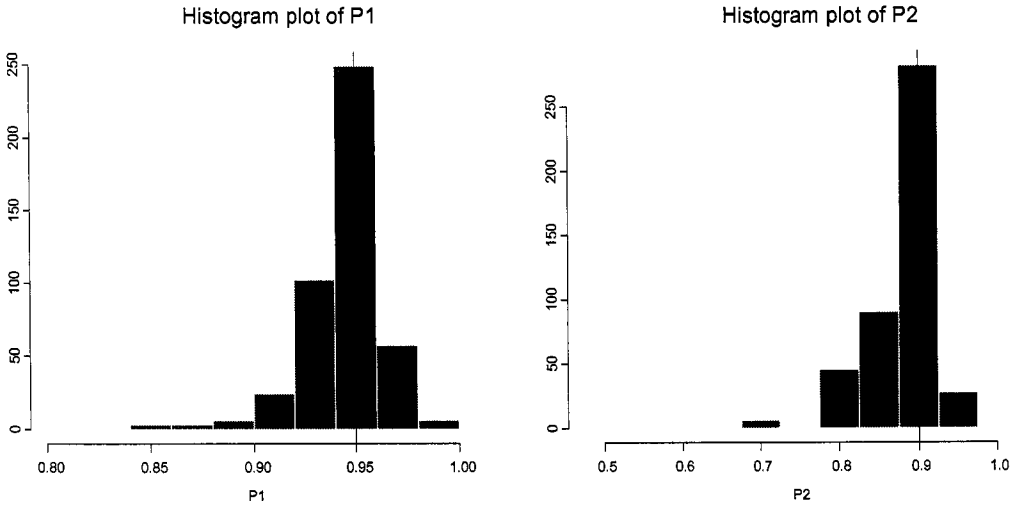


Figure 3.5a: Distribution of EMM Estimates (T=4000) I

The distributions of EMM estimates $(\mu, \beta, \sigma_1, \sigma_2)$ are plotted using the BIC-preferred ARCH-leading SNP model 10516000 (solid line) and GARCH-leading model 11116000 (dotted line) with vertical lines as the associated true values.

(A) ARCH-leading SNP



(B) GARCH-leading SNP

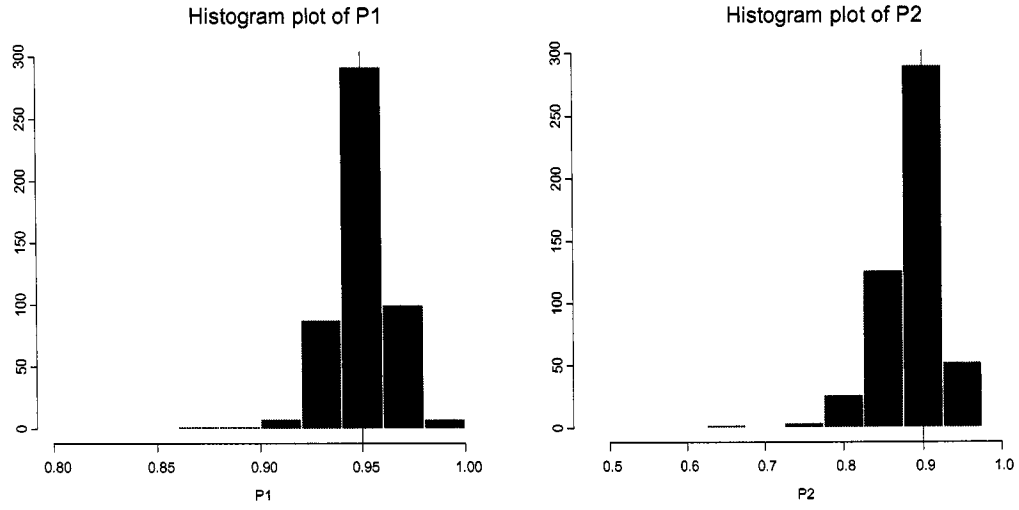


Figure 3.5b: Distribution of EMM Estimates (T=4000) II

The histograms of EMM estimates (P_1, P_2) are plotted using the BIC-preferred ARCH-leading SNP model 10516000 (Panel A) and GARCH-leading model 11116000 (Panel B) with vertical lines as the associated true values.

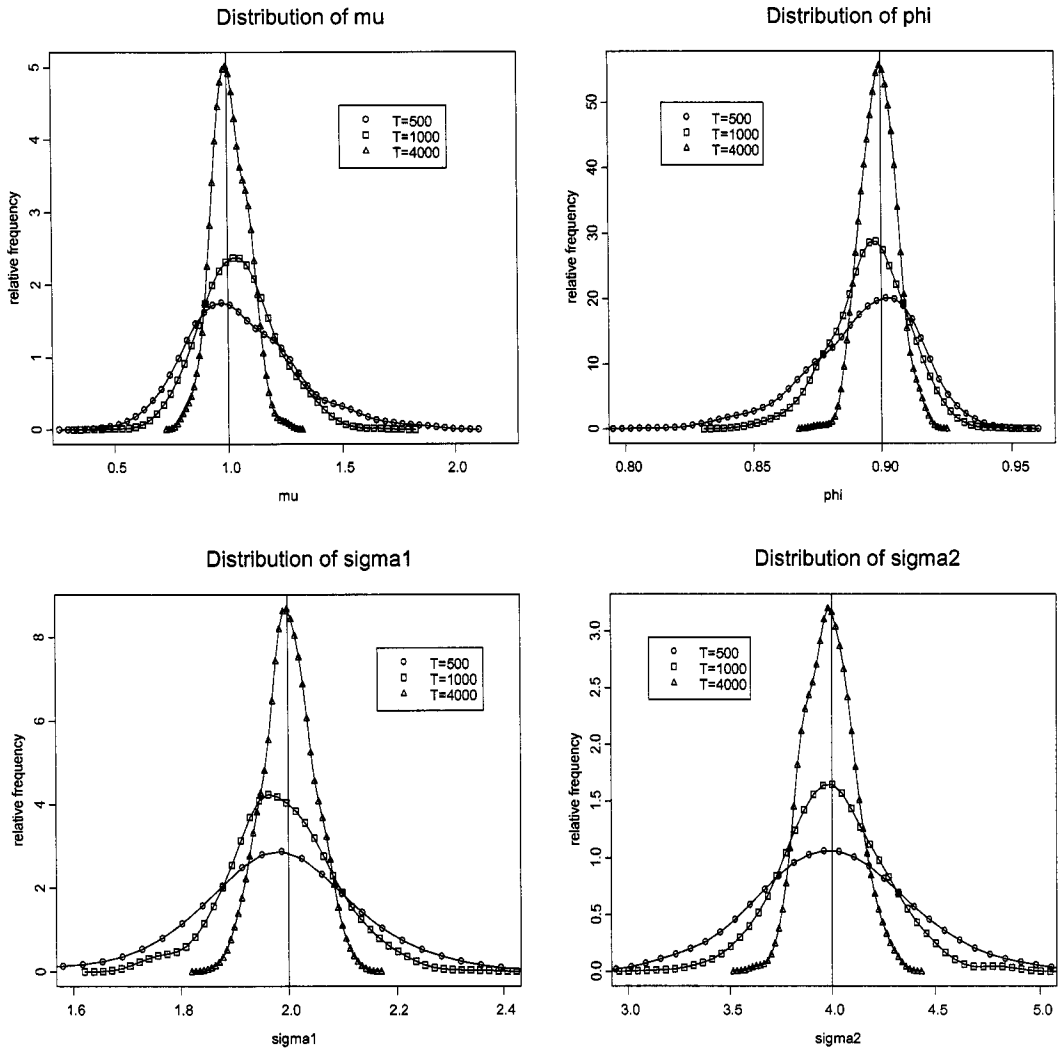


Figure 3.6a: Distribution of State Space ML Estimates I

The distributions of State-Space ML estimates $(\mu, \beta, \sigma_1, \sigma_2)$ are plotted for sample sizes of T=500, T=1000, and T=4000 with vertical lines as the associated true values.

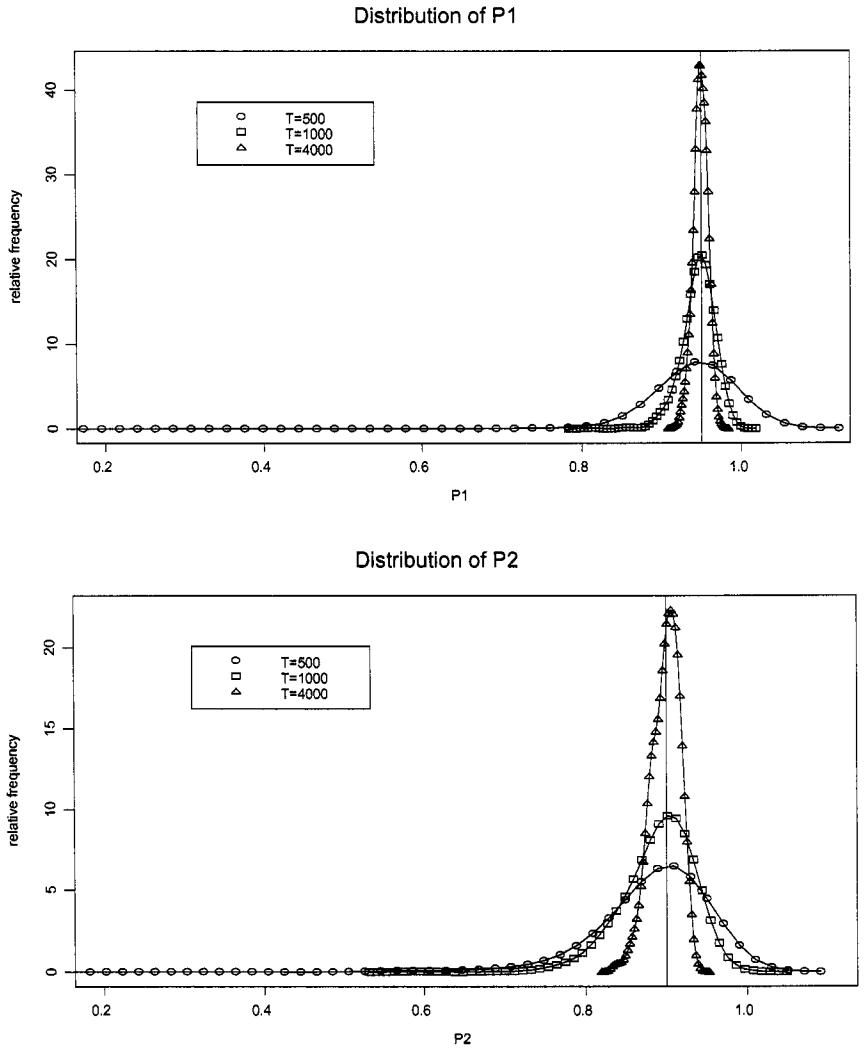


Figure 3.6b: Distribution of State Space ML Estimates II

The distributions of State-Space ML estimates (P_1, P_2) are plotted for sample sizes of T=500, T=1000, and T=4000 with vertical lines as the associated true values.

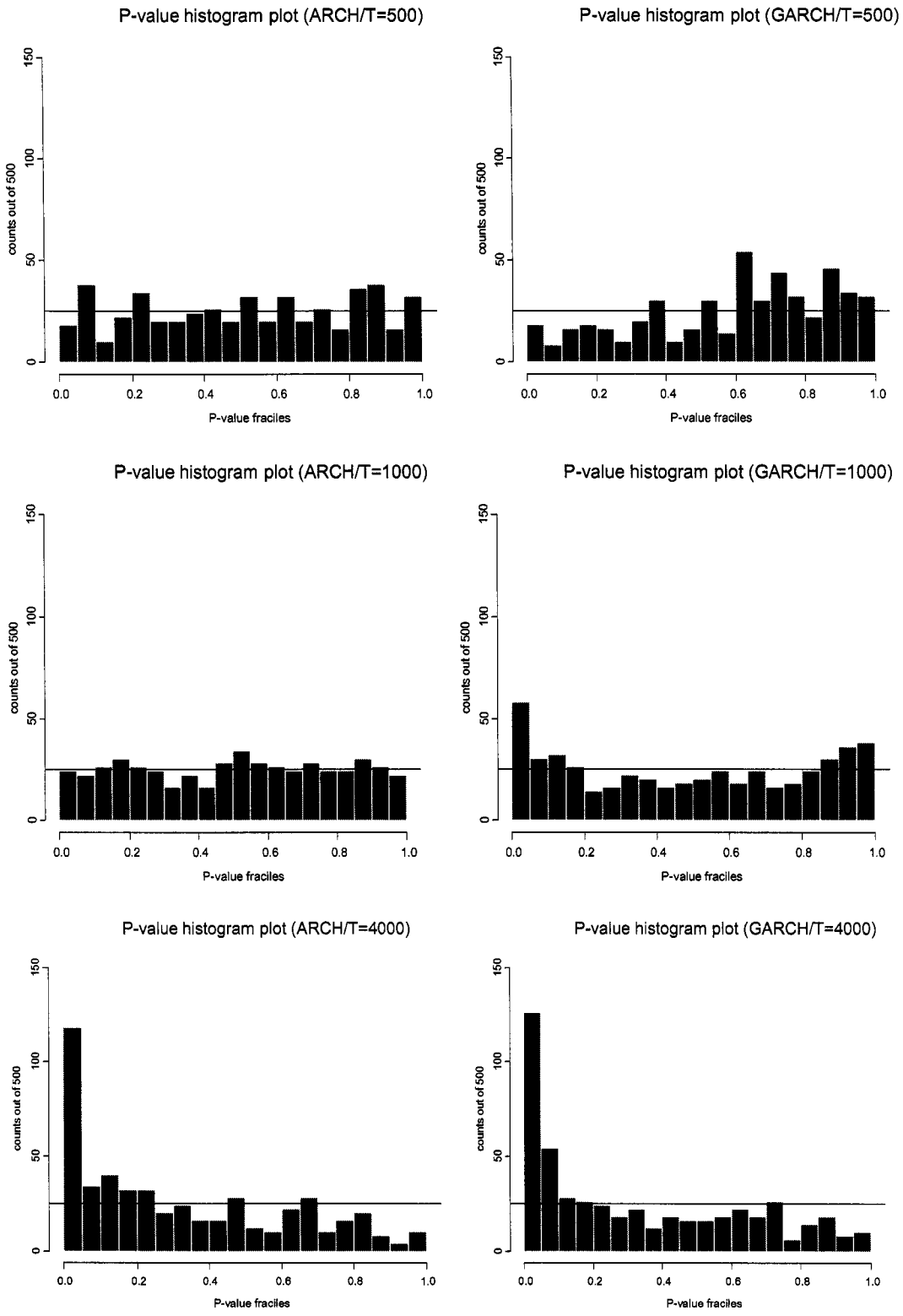


Figure 3.7: Histogram for EMM P-values

Table 3.1: Statistics Summary of Simulated Data

The statistics summary is given in panel (A) for the simulated data from the RS model (3.4) with 1000 observations. The panel (B) shows the statistics summary for the change of the simulation.

(A)

Sample Quantiles	Min: -15.08	1Q: 5.73	Median: 9.99	3Q: 13.27	Max:32.73
Sample Moments	Mean: 9.65	Std. Dev.: 6.22	Skewness: -0.00	Kurtosis: 3.53	

(B)

Sample Quantiles	Min: -14.25	1Q: -1.76	Median: 0.004	3Q: 1.64	Max:14.21
Sample Moments	Mean:0.02	Std. Dev. : 2.86	Skewness: -0.04	Kurtosis: 4.87	

Table 3.2: Monte Carlo Results (T=500)

The mean, 5% quantile, 95% quantile, and associated root mean squared errors (RMSEs) of the EMM estimates (with p-value) are reported in top panel for the BIC-preferred ARCH leading score generator (10210000) and in middle panel for the BIC-preferred GARCH leading one (11110000). The same information of State Space ML estimates using Kim's filter is provided in the bottom panel.

T=500	μ	β	σ_1	σ_2	P_1	P_2
True Values	**1**	**0.9**	**2**	**4**	**0.95**	**0.90**
EMM ARCH-Leading 10210000						
5% quantile	0.6677	0.8536	1.7678	3.8719	0.8300	0.5500
Mean	1.0653	0.8931	1.9662	4.1659	0.9262	0.8205
95% quantile	1.4927	0.9247	2.1581	4.6942	0.9700	0.9500
RMSE	0.2718	0.0219	**0.1288**	**0.2775**	0.0470	0.1271
EMM GARCH-Leading 11110000						
5% quantile	0.6572	0.8548	1.7569	3.8759	0.8700	0.6000
Mean	1.0564	0.8938	1.9323	4.1824	0.9282	0.8262
95% quantile	1.4952	0.9261	2.1089	4.6842	0.9700	0.9500
RMSE	**0.2603**	**0.0212**	0.1622	0.3089	**0.0366**	**0.1021**
State Space Model						
5% quantile	0.7223	0.8539	1.7586	3.4326	0.8859	0.7639
Mean	1.0612	0.8932	1.9800	4.0079	0.9398	0.8804
95% quantile	1.5127	0.9204	2.2177	4.6218	0.9763	0.9584
RMSE	**0.2377**	**0.0207**	0.1437	0.3703	**0.0343**	**0.0770**

Table 3.3: Monte Carlo Results (T=1000)

The mean, 5% quantile, 95% quantile, and associated root mean squared errors (RMSEs) of the EMM estimates (with p-value) are reported in top panel for ARCH leading score generator (10310000) and in middle panel for GARCH leading score generator (11114000). The same information of State Space ML estimates using Kim's filter is provided in the bottom panel.

T=1000	μ	β	σ_1	σ_2	P_1	P_2
True Values	**1**	**0.9**	**2**	**4**	**0.95**	**0.90**
EMM ARCH-Leading 10310000						
5% quantile	0.7514	0.8689	1.8455	3.8408	0.8900	0.7500
Mean	1.0385	0.8960	1.9843	4.0816	0.9431	0.8648
95% quantile	1.3569	0.9206	2.1149	4.4152	0.9700	0.9500
RMSE	0.1874	0.0160	**0.0819**	**0.1823**	0.0287	0.0683
EMM GARCH-Leading 11114000						
5% quantile	0.7857	0.8683	1.8865	3.6983	0.8900	0.7500
Mean	1.0501	0.8953	2.0159	3.9705	0.9407	0.8554
95% quantile	1.3403	0.9187	2.1650	4.3261	0.9700	0.9000
RMSE	**0.1754**	**0.0156**	0.0832	0.2192	**0.0284**	**0.0631**
State Space Model						
5% quantile	0.7959	0.8713	1.8384	3.5954	0.9120	0.8149
Mean	1.0470	0.8956	1.9881	4.0048	0.9455	0.8888
95% quantile	1.3329	0.9176	2.1524	4.4196	0.9745	0.9420
RMSE	**0.1632**	**0.0144**	0.0971	0.2490	**0.0201**	**0.0431**

Table 3.4: Monte Carlo Results (T=4000)

The mean, 5% quantile, 95% quantile, and associated root mean squared errors (RMSEs) of the EMM estimates (with p-value) are reported in top panel for ARCH leading score generator (10516000) and in middle panel for GARCH leading score generator (11116000). The same information of State Space ML estimates using Kim's filter is provided in the bottom panel.

T=4000	μ	β	σ_1	σ_2	P_1	P_2
True Values	**1**	**0.9**	**2**	**4**	**0.95**	**0.90**
EMM ARCH-Leading 10516000						
5% quantile	0.8690	0.8866	1.9269	3.7269	0.9100	0.8000
Mean	1.0131	0.8991	2.0176	3.9469	0.9444	0.8803
95% quantile	1.1692	0.9122	2.1111	4.1994	0.9700	0.9500
RMSE	**0.0902**	**0.0075**	0.0641	0.1922	0.0193	0.0424
EMM GARCH-Leading 11116000						
5% quantile	0.8716	0.8716	1.9210	3.6782	0.9300	0.8000
Mean	1.0151	0.8988	2.0088	3.9565	0.9499	0.8852
95% quantile	1.1584	0.9121	2.0987	4.2297	0.9700	0.9500
RMSE	0.0909	0.0078	**0.0552**	**0.1838**	**0.0155**	**0.0403**
State Space Model						
5% quantile	0.8844	0.8876	1.9229	3.8065	0.9332	0.8688
Mean	1.0109	0.8991	2.0004	3.9870	0.9489	0.8987
95% quantile	1.1363	0.9098	2.0768	4.1878	0.9625	0.9240
RMSE	**0.0778**	**0.0068**	**0.0466**	**0.1194**	**0.0090**	**0.0176**

Table 3.5: EMM Estimation Based on SV Model

For the simulated data (T=1000) from the RS model (3.4), the mean, 5% quantile, and 95% quantile of the EMM estimations are given for the misspecified two-factor SV model by (3.6). The BIC preferred ARCH-leading (10310000) and GARCH-leading (11114000) SNP models are used as the auxiliary models.

True Model	μ	β	σ_1	σ_2	P_1	P_2	
Parameters	**1**	**0.9**	**2**	**4**	**0.95**	**0.90**	
Misspecified Model	μ	β	ω_0	ω_1	ξ	p-value	Rejection rate (5%)
EMM ARCH-Leading 10310000							
5% quantile	43.36	4.53	0.19	0.03	0.00	0.00	
Mean	80.98	7.55	62.81	9.86	1.47	20.65	68.33%
95% quantile	93.47	13.67	152.72	27.77	5.95	75.33	
EMM ARCH-Leading 11114000							
5% quantile	44.23	4.43	0.58	0.10	0.00	0.00	
Mean	67.38	7.40	62.04	10.56	2.52	28.13	76.52%
95% quantile	76.47	10.38	151.00	26.47	5.65	84.77	

REFERENCES

Ahn, D.-H., Dittmar, R.F., Gallant, A.R., and Gao, B., 2003. "Purebred or Hybrid? Reproducing the volatility in term structure dynamics", Journal of Econometrics 116, 147-180.

Ahn, D.-H., Dittmar, R.F., Gallant, A.R., and Gallant, A.R., 2002, "Quadratic Term Structure Models: Theory and Evidence," Review of Financial Studies 15(1), 243-88.

Andersen, T.G., Benzoni, L., and Lund, J., 2002. "An Empirical Investigation of Continuous-Time Models for Equity Returns," Journal of Finance 57, 1239-1284.

Andersen, T.G., and Bollerslev, T., 2005. "A Framework for Exploring the Macroeconomic Determinants of Systematic Risk", Working Paper, Annual Meeting of American Economics Association.

Andersen, T.G., Chung, Hyung-Jin, and Sorensen, B.E., 1999. "Efficient Method of Moments Estimation of a Stochastic Volatility Model: A Monte Carlo Study", Journal of Econometrics, 91, 61-87.

Andersen, T.G. and Sorensen, B.E., 1996. "GMM Estimation of A Stochastic Volatility Model: A Monte Carlo Study", Journal of Business and Economic Statistics, 14, 328-352.

Andersen, T.G. and Sorensen, B.E., 1997. "Estimating Continuous-time Stochastic Volatility Models of the Short-term Interest Rate", Journal of Business and Economic Statistics, 14, 328-352.

Ang, A., Bekaert, G., 2002. "Regime Switching in Interest Rates", Journal of Business and Economic Statistics 20(2), 163-182.

Bansal, R. and Zhou, H., 2002. "Term structure of interest rates with regime shifts", The Journal of Finance LVII(5) , 1997-2043.

Barndorff-Nielsen, O.E., 1997. "Normal Inverse Gaussian Distributions and Stochastic Volatility Modeling", Scand Journal of Statistics 24, 1-13.

Bergman, N. "Recursive Bayesian Estimation: Navigation and Tracking Applications", Ph.D. dissertation, Linkoping University, Linkoping, Sweden. 1999.

Bertoin, J., 1996. "Lèvy Processes", Cambridge University Press, Melbourne, NY.

Bliss, R.R. and Smith, D.C. 1998. "The Elasticity of Interest Rate Volatility: Chan, Karolyi, Longstaff, and Sanders Revisited", Journal of Risk 1(1), 21-246.

Bollerslev, T., 1986. "Generalized Autoregressive Conditional Heteroskedasticity", Journal of Econometrics 31, 307-327.

Boudoukh, J., Richardson, M., Smith, T., and Whitelaw, R., 1999. "Regime Shifts and Bond Returns", working paper, New York University.

Brandt, M. and Chapman, D., 2002. "Comparing Multifactor Models of the Term Structure", working paper, Duke University.

Broto, C. and Ruiz, E., 2002. "Estimation Methods for Stochastic Volatility Models: A Survey", Journal of Economic Surveys 18(5), 613-649.

Cai, J., 1994. "A Markov Model of Switching-Regime ARCH". Journal of Business and Economic Statistics 12, 309-316.

Chan, K.C., Karolyi, G.A., Longstaff, F., and Sanders, A. 1992. "The Volatility of Short-term Interest Rates: An empirical Comparison of Alternative Models of the Term Structure of Interest Rates", Journal of Finance 47, 1209-1227.

Carpenter, J., Clifford, P., and Fearnhead, P., 1999. "Improved Particle Filter for Nonlinear Problems", Proc. Inst. Elec. Eng., Radar, Sonar, Navig.

Carr, P.,Geman, H., Madan, D., and Yor, M., 2000. "The Fine Structure of Asset Returns: An Empirical Investigation", Journal of Business 75(2), 305-332.

Chernov, M. and Ghysels, E., 2000. "A Study towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation," Journal of Financial Economics 56, 407-458.

Chernov, M., Gallant, A.R., Ghysels, E., and Tauchen, G., 2003. "Alternative Models for Stock Price Dynamics", *Journal of Econometrics*, forthcoming.

Christiansen, C., 2005. "Level-ARCH Short Rate Models with Regime Switching: Bivariate Modeling of US and European Short Rates", Working Paper, Centre for Analytical Finance (CAF).

Cox, J.C., Ingersoll, J.E., and Ross, S.A., 1985. "A Theory of the Term Structure of Interest Rates", Econometrica 53, 385-407.

Cosslett, S.R., and Lee, L.F., 1985. "Serial Correlation in Latent Discrete Variable Models", Journal of Econometrics, 27, 79-97.

Dahlquist, M. and Gray, S.F., 2000. "Regime-Switching and Interest Rates in the European Monetary System", Journal of International Economics 50, 399-419.

Dai, Q. and Singleton, K.J., 2003. "Term Structure Dynamics in Theory and Reality", Review of Financial Studies.

Dai, Q. and Singleton, K.J., 2000. "Specification Analysis of Affine Term Structure Models," Journal of Finance 55, 1943-1978.

Dai, Q., Singleton, K.J., and Yang, W., 2004. "Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields," Working paper, New York University.

Danielsson, J., 1994. "Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood", Journal of Econometrics 61, 375–400

Davies, R.B., 1977. "Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative", *Biometrika*, 64, 247-254.

Davies, R.B., 1987. "Hypothesis Testing When a Parameter Is Present Only under the Alternative", *Biometrika,* **74,** 33-43.

Diebold, F.X. and Glenn, D.R., 1996. "Measuring Business Cycles: Modern Perspective", Review of Economics and Statistics, 78, 67-77.

Doucet, A., Godsill, S., and Andrieu C., 2000. "On Sequential Monte Carlo Methods for Bayesian Filtering", Statistics and Computing 10, 197-208.

Driffill, J., T. Kenc, M. Sola, and F. Spagnolo (2004): "An Empirical Examination of Term Structure Models with Regime Shifts," Discussion paper, Centre for Economic Policy Research.

Duffee, G., 1993. "On the Relation between the Level and Volatility of Short-Term Interest Rates: A Comment on Chan, Karolyi, Longstaff and Sanders", Working Paper, Federal Researve Board Washington D.C.

Duffie, D. and Singleton, K., 1989. "Simulated Moments Estimation of Markov Models of Asset Prices", working paper, Graduate School of Business, Stanford Unversity.

Duffie, D., Pan, J., and Singleton, K., 2000. "Transform Analysis and Asset Pricing for Affine Jump Diffusions", Econmetrica 68, 1343-1376.

Engle, R.F., 1982. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation", Econometrica 50, 987-1007.

Engle, R.F. and Gonzalez-Rivera, G., 1991. "Semiparametric ARCH Models", Journal of Business and Economic Statistics 9(4), 345-359.

Engle, R.F. and Hamilton, J.D., 1990. "Long Swings in the Dollar: Are They in the Data and Do Markets Know It?" American Economic Review, 80, 689-713.

Evans, M., 2001. "Real Risk, Inflation Risk, and the Term Structure", working paper, Georgetown University.

Fama, E.F., 1965. "The Behavior of Stock Market Prices", Journal of Business 38, 34-105.

Filardo, A.J., 1994. "Business Cycle Phases and Their Transitional Dynamics", Journal of Business and Economic Statistics, 12, 299-308.

Fridman, M. and Harris, L., 1998. "A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models", Journal of Business and Economic Statistics, 16, 284-291.

Gallant, A. R, Hsieh D.A., and Tauchen, G., 1997. "Estimation of Stochastic Volatility Models with Diagnostics," Journal of Econometrics, 81(1), 159-192.

Gallant, R.A. and Tauchen, G., 2001. "Efficient method of moments", Manuscript, University of North Carolina.

Gallant, A.R. and Tauchen, G., 1998. "Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions", Journal of American Statistical Association, 93(441), 10-24.

Gallant, A.R. and Tauchen, G., 2001. "Efficient method of moments", Manuscript, University of North Carolina.

Gallant, R.A. and Tauchen, G., 2001. "SNP: A program for nonparametric time series analysis, Version 8.8, User's Guide", Manuscript, University of North Carolina.

Gallant, R.A. and Tauchen, G., 2002. "EMM: A program for efficient method of moments estimation, Version 1.6, User's Guide", Manuscript, University of North Carolina.

Garcia, P. and Perron, P., 1996. "An Analysis of the Real Interest Rate Under Regime Shifts", Review of Economic and Statistics 78(1), 111-125.

Garcia, R., Renault, E. and Veredas, D., 2004. "Estimation of Stable Distributions by Indirect Inference", working paper.

Geweke, J., 1994. "Bayesian Comparison of Econometric Models", working paper, Federal Reserve Bank of Minneapolis, Minnesota.

Gordon, N., Salmond, D., and Smith, A.F.M., 1993. "Novel Approach to Non-linear and Non-Gaussian Bayesian State Estimation", Proc. Inst. Elect. Eng., F, 140, 107-113.

Gray, S.F., 1996. "Modeling the conditional distribution of interest rate as a regime-switching process", Journal of Financial Economics 42, 27-62.

Hamilton, J.D., 1988. "Rational Expectations Econometric Analysis of Changes in Regimes: An Investigation of the Term Structure of Interest Rates", Journal of Economic Dynamics and Control, 12, 385-432.

Hamilton, J.D., 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", Econometrica, 57(2), 357-384.

Hamilton, J.D., 1990. "Analysis of Time Series Subject to Changes in Regime", Journal of Econometrics, 45, 39-70.

Hamilton, J.D., 1993. "Estimation, Inference, and Forecasting of Times-Series Subject to Changes in Regime", Handbook of Statistics, 11, 231-259. New York: Elevie Science Publisher B.V.

Hamilton, J.D., 1994. "Time Series Analysis", Princeton University Press, Princeton.

Hamilton, J.D. and Susmel, R., 1994. "Autoregressive Conditional Heteroskedasticity and Changes in Regime", Journal of Econometrics 64, 307-333.

Harrison, P.J., and Stevens, C.F., 1976. "Bayesian Forecasting", Journal of the Royal Statistical Society, Series B, 38, 205-247.

Harvey, A., Ruiz, E., and Shephard, N, 1994. "Multivariate Stochastic Variance Models", Review of Economic Studies, Blackwell Publishing, 61(2), 247-264.

Harvey, A. and Shephard, N., 1996. "Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns", Journal of Business and Economic Statistics 14, 429-434.

Iorio, F.D. and Calzolari, G. 2005. "Discontinuities in Indirect Estimation: An Application to EAR Models", Computational Statistics and Data Analysis, forthcoming.

Jacquier, E., Polson, N.G., and Rossi, P.E., 1994. "Bayesian Analysis of Stochastic Volatility Models", Journal of Business and Economic Statistics 12(4), 413-17.

Javaheri, A., Lautier, D., and Galli, A., 2003. "Filtering in Finance", Wilmott Magazine, issue 5.

Jondeau E. and Rockinger M., 2003. "Conditional volatility, Skewness, and Kurtosis: Existence, Persistence, and Comovements", Journal of Economic Dynamics & Control, 27, 1699-1737.

Kim, Chang-Jin, 1994, "Dynamic Linear Models with Markov-Switching", Journal of Econometrics 60, 1-22.

Kim, Chang-Jin, and Nelson, C.R., 1999, "State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications", MIT Press.

Lahiri, K. and Wang, J.G., 1996. "Interest Rate Spreads as Predictors of Business Cycles", Statistical Methods in Finance: Handbook of Statistics (14), Edited by G.S.Maddala and C.R.Rao, North Holland, 297-315.

Lahiri, K., Ivanova, D. and Seitz, F., 2000. "Interest Rate Spreads as Predictors of German Inflation and Business Cycle", International Journal of Forecasting, 39-58.

Lambert, P. and Laurent S. 2000. "Modeling financial time series using GARCH-type models with a skewed student distribution for the innovations", Discussion Paper 0125, Institute de Statistique, Universite Catholique de Louvain.

Landen, C., 2000. "Bond Pricing in a Hidden Markov Model of the Short Rate", Finance and Stochastics 4, 371-389.

Liechty, J. C., and Roberts, G.O., 2001. "Markov Chain Monte Carlo Methods for Switching Diffusion Models," Biometrika, 88(2), 299-315.

Liesenfeld, R. and Richard, J.F. 2003. "Univariate and multivariate stochastic volatility models: estimation and diagnostics", Journal of Empirical Finance, 10, 505-531.

Liu, S-Mi. and Brorsen, B.W. 1995. "Maximum Likelihood Estimation of a GARCH-Stable Model", Journal of Applied Econometrics, 10, 275-285.

Liu, M., 2000. "Modeling Long Memory in Stock Market Volatility," Journal of Econometrics 99(1): 139-71.

Lo, A.W., 1988. "Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data," Econometric Theory 4, 231-247.

Longstaff, F.A. and Schwartz, E.S., 1992. "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model", Journal of Finance 47(4), 1259-1282.

Matacz, A., 2004. "Financial Modeling and Option Theory with the Truncated Lèvy Process", working paper.

McAleer, M., 1995. "The Significance of Testing Empirical Nonnested Models", Journal of Econometrics, 67, 149-171.

Melino, M. and Turnbull, S.M., 1990. "Pricing Foreign Currency Options with Stochastic Volatility", Journal of Econometrics 45, 239-265.

Marinelli C. and Rachev, S., 2002. "Some Applications of Stable Models in Finance", working paper.

Merton, R., 1976. "Option Pricing When Underlying Stock Returns are Discontinuous", Journal of Financial Economics 3, 125-144.

Naik, V. and Lee, M. H., 1997. "Yield Curve Dynamics with Discrete Shifts in Economic Regimes: Theory and Estimation", Working paper, University of British Columbia.

Pesaran, B, and Pesaran, M.H., 1993. "A Simulation Approach to the Problem of Computing Cox's Statistic for Testing Non-nested Models", *Journal of Econometrics*, 57, 377-392.

Pesarn, M., and Potter, M., 1992. "Non-linear Dynamics and Econometrics: An Introduction", *Journal of Applied Econometrics* (Special Issue), 7, S1-S7.

Pitt M. and Shephard, N., 1999. "Filtering via Simulation: auxiliary Particle Filters", Journal of American Statistical Association, 94(446), 590-599.

Ripley, B., 1987. "Stochastic Simulation", Wiley, New York.

Sandmann, G. and Koopman, S.J., 1998, "Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood", Journal of Econometrics, 87(2), 271-301.

Smith, D.R., 2002. "Markov-Switching and Stochastic Volatility Diffusion Models of Short-Term Interest Rates", Journal of Business and Economic Statistics 20(2) 183-97.

So, M.K., Lam, K. and Li, W.K., 1998. "A Stochastic Volatility Model with Markov Switching", Journal of Business and Economic Statistics 7, 177-192.

Sola, M. and Driffill, J., 1994. "Testing the Term Structure of Interest Rates Using a Stationary Vector Autoregression with Regime Switching", Journal of Economic Dynamics and Control, 18, 601–628.

Taylor, S.J., 1986. "Modeling Financial Time Series", Wiley, New York.

Taylor, S.J., 1994. "Modeling Stochastic Volatility", Mathematical Finance 4, 183-204.

Timmermann, A., 2000. "Moments of Markov Switching Models", Journal of Econometrics, 96, 75–111.

Wu, S. and Zeng, Y., 2003. "Regime-switching Risk in the Term Structure of Interest Rates", working paper, University of Kansas.

Zivot, E., and Wang, J., 2005. "Modeling Financial Time Series with S-PLUS", Springer-Verlag Publishing, forthcoming.

Appendix A: Implementation of SNP Selection

To enhance the searching efficiency, we utilize the following strategies for determining the most appropriate SNP model. The expository discussion is in Gallant and Tauchen (2001).

(1) As a general rule with financial data, we always move K_z from 0 up to 4. Due to the fat-tailed error densities relative to the Gaussian for financial data, the polynomials has to increase the mass around zero, depress the mass on either side of zero and then increase the mass in the tails by going to infinity on the left and right side. Not linear, quadratic, cubic, but only the quartic polynomial is needed in order to reach the above goal easily and successfully. (2) We put an upper bound of 8 for K_z in order to improve the stability of computation, because the polynomials fit little wiggles when $K_z > 8$. (3) We also put an upper bound of 8 for L_r when fitting the SNP density as VAR-ARGH leading terms. (4) The spline transformation is recommended to use, which is essential for extremely persistent data such as interest rates. (5) In processing a specific starting parameter set, we perturb each active parameter as

$$\rho_i \to (1 + u \times tweak)\rho_i$$

where u is uniform (-1,1), then iterate from these values for 10 iterations, and repeat this process for many trials. Lastly, it iterates from the best parameter values of these 10 trials until convergence. Therefore, bad starting values leading to local optima are not a concern. This random restart strategy yielded satisfactory fits, sometime improving the estimations substantially; we also utilize this strategy in the estimation step.

Appendix B: Kim's Approximation Algorithm for Markov RS Model

A state space model can be represented as follows:

$$\begin{pmatrix} \alpha_{t+1} \\ y_t \end{pmatrix} = \delta_t + \Phi_t \alpha_t + u_t \tag{B.1}$$

$$\delta_t = \begin{pmatrix} d_t \\ c_t \end{pmatrix}, \Phi_t = \begin{pmatrix} T_t \\ Z_t \end{pmatrix}, u_t = \begin{pmatrix} H_t \eta_t \\ G_t \varepsilon_t \end{pmatrix}, \Omega_t = \begin{pmatrix} H_t H_t' & 0 \\ 0 & G_t G_t' \end{pmatrix}$$

where α_{t+1} is the $M \times 1$ state vector, y_t is the $N \times 1$ vector of observed variables, $\eta_t \sim iidN(0, I_r)$ is the $r \times 1$ vector of disturbance terms in the transition equation governing the dynamics of the state vector α_{t+1}, $\varepsilon_t \sim iidN(0, I_r)$ is the $N \times 1$ vector of disturbance terms in the measurement equation governing the dynamics of the observed variables y_t, and d_t, T_t, H_t, c_t, Z_t and G_t are conformable parameter matrices or system matrices.

For Markov switching state space models, the parameter matrices are assumed to be dependent on a latent or unobserved discrete state variable S_t:

$$\delta_t = \delta_{S_t}, \Phi_t = \Phi_{S_t}, \Omega_t = \Omega_{S_t} \tag{B.2}$$

and the discrete state variable S_t follows a *k*-regime Markov chain given by

$$\begin{aligned} p_{ij} &= \Pr(S_t = j \mid S_{t-1} = i) \\ \sum_{j=1}^{k} p_{ij} &= 1 \end{aligned} \tag{B.3}$$

For this paper, by setting the continuous state vector to $\alpha_{t+1} = (y_{t+1}, y_t)$, the RS model (3.4) can be put into the above state space representation with:

$$\delta_t = \begin{pmatrix} \mu \\ 0 \\ 0 \end{pmatrix}, \Phi_t = \begin{pmatrix} \beta & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix}, \Omega_t = \begin{pmatrix} \sigma_{S_t}^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{B.4}$$

Since Markov switching state space models allow for nonlinear dynamics, the traditional Kalman filtering and smoothing algorithms for Gaussian linear state space models can no longer be applied to obtain valid inference on the unobserved state vector.

In particular, given the initial estimate, the prediction equations of the Kalman filter for Gaussian linear state space model become:

$$a_{t+1|t}^{(i,j)} = T_t a_{t|t}^{(i)} \tag{B.5}$$

$$P_{t+1|t}^{(i,j)} = T_t P_{t|t}^{(i)} T_t' + H_t H_t' \tag{B.6}$$

where $a_{t|t} = E[\alpha_t \mid Y_t]$, $P_{t|t} = \mathrm{var}[\alpha_t \mid Y_t]$ with the superscript (i, j) that denotes the case of $S_t = i$ and $S_{t+1} = j$ for $i, j = 1,...,k$. The updating equations of the Kalman filter for Gaussian linear state space models become:

$$a_{t+1|t}^{(i,j)} = a_{t|t-1}^{(i,j)} + K_t^{i,j} v_t^{(i,j)} \tag{B.7}$$

$$P_{t|t}^{(i,j)} = P_{t|t-1}^{(i,j)} - P_{t|t-1}^{(i,j)} Z_t' (K_t^{(i,j)})' \tag{B.8}$$

where

$$\begin{cases} v_t^{(i,j)} = y_t - c_t - Z_t a_{t|t-1}^{(i,j)} \\ K_t^{(i,j)} = P_{t|1}^{(i,j)} Z_t' (F_t^{(i,j)})^{-1} \\ F_t^{(i,j)} = Z_t P_{t|t-1}^{(i,j)} Z_t' + G_t G_t' \end{cases} \tag{B.9}$$

Obviously even for a relatively small sample, the Kalman filtering algorithm will become computationally infeasible. To make the filtering algorithm manageable, Kim (1994) proposes to collapse the set of statistics in the updating equations (3.14)-(3.15) as follows:

$$a_{t|t}^{(j)} = \frac{\sum_{i=1}^{k} \Pr(S_t = j, S_{t-1} = i \mid Y_t) a_{t|t}^{(i,j)}}{\Pr(S_t = j \mid Y_t)} \tag{B.10}$$

$$P_{t|t}^{(j)} = \frac{\sum_{i=1}^{k} \Pr(S_t = j, S_{t-1} = i \mid Y_t)[P_{t|t}^{(i,j)} + (a_{t|t}^{(j)} - a_{t|t}^{(i,j)})(a_{t|t}^{(j)} - a_{t|t}^{(i,j)})']}{\Pr(S_t = j \mid Y_t)} \tag{B.11}$$

where the filtered probability $\Pr(S_t = j \mid Y_t)$ can be updated given an initial estimate. Now at each step, only k sets of statistics need to be stored, which can be fed into the prediction equations (B.7)-(B.8) to complete the filtering algorithm. Moreover, the log likelihood function can be obtained from

$$L(.) = \sum \log f(y_t \mid I_{t-1}) = \sum \log \left\{ \sum_{j=1}^{k} f(y_t \mid I_{t-1}, S_t = j) \Pr(S_t = j \mid I_{t-1}) \right\} \tag{B.12}$$

which can be maximized with respect to unknown parameters of the model.

VITA

Ying Gu was born in Beijing, China. She earned a Bachelor of Arts degree in Economics at Beijing University, in Beijing, China before beginning graduate studies in economics at the University of Washington, Seattle, in 2001. She received a Master of Arts degree in Economics and a Graduate Certificate in Computational Finance from the University of Washington in 2004 and in 2006, respectively.