Evaluating structural models for the U.S. short rate using EMM and optimal filters

Drew Creal, Ying Gu, and Eric Zivot*

First version: August 10, 2006 Current version: March 17, 2007

Abstract

We combine the efficient method of moments with appropriate algorithms from the optimal filtering and signal extraction literature to study a collection of models for the U.S. short rate. Our models include two continuous time stochastic volatility models and two regime switching models, which provided the best fit in previous work that examined a larger collection of models. The continuous time stochastic volatility models fall into the class of nonlinear, non-Gaussian state space models for which we apply particle filtering and smoothing algorithms. Our results demonstrate the effectiveness of the particle filter for continuous time processes. Our analysis also provides an alternative and complementary approach to the reprojection technique of Gallant and Tauchen (1998) for studying the dynamics of volatility.

*PRELIMINARY. Comments are welcome. Authors' contact: Drew Creal, Department of Econometrics, Free University, Amsterdam, The Netherlands, dcreal@feweb.vu.nl; Ying Gu, Bear Stearns, New York, yinggu@u.washington.edu; Eric Zivot, ezivot@u.washington.edu, Department of Economics, University of Washington, Seattle. The first author would like to gratefully acknowledge the financial support of the Grover and Creta Ensley Fellowship while at the University of Washington. All remaining errors are our own.

1 Introduction

The efficient method of moments (EMM) of Gallant and Tauchen (1996) has proven to be a flexible method for estimating structural time series models. A major advantage of EMM is its ease of implementation relative to other methodologies. This is particularly true for more challenging models in finance such as discrete and continuous time stochastic volatility models. While there is a sizeable literature that uses EMM to estimate parameters of volatility models, little attention has been paid within this literature to estimating the latent volatility after estimating the parameters.

To solve this problem, Gallant and Tauchen (1998) added a reprojection step to EMM that computes the conditional volatility of the one-step ahead predictive density of the time series. Instead of adopting this strategy, we combine EMM with appropriate algorithms from the literature on optimal filtering and signal extraction.[1] Specifically, we first estimate the parameters of candidate volatility models for the U.S. short rate using EMM. With the parameter estimates in hand, we then use optimal filtering and smoothing algorithms to infer the unobserved state variables that include the volatilities. We show how the one-step ahead predicted, filtered and smoothed estimates can be used to compare and evaluate models.

We consider four models from Gu and Zivot (2006), who performed an exhaustive comparison of univariate models for the U.S. short rate. In their study, a two-factor and a three-factor stochastic volatility (SV) model and several regime-switching (RS) models passed the EMM specification test at a 5% significance level. The continuous time SV models specify returns as a nonlinear function of volatility. We apply particle filtering and smoothing algorithms to these models. Particle filters were introduced by Gordon *et. al.* (1993) to estimate state variables in nonlinear, non-Gaussian state space models. They were first considered for discrete time SV models by Kim *et. al.* (1998) and more recently for diffusion-driven models by Johannes *et. al.* (2006) and Fearnhead *et. al.* (2006). The second set of models we consider are regime switching models whose latent states are discrete valued. In this case, the optimal filter is the discrete state hidden Markov model (HMM) filter first developed by Baum and coauthors, see Baum and Petrie (1966) and Baum *et. al.* (1970).

The importance of recovering the latent volatility through filtering and smoothing is two-fold. First, forecasting future volatility is critical for asset pricing, portfolio management, and risk man-

[1] Chib *et. al.* (2002) also proposed using particle filters as an alternative to Gallant and Tauchen's reprojection step for discrete time SV models. In this paper, we also consider particle smoothing, maximum a posteriori sequence estimation, as well as discrete-state methods.

agement. Filtering algorithms provide a convenient method to make forecasts because the one-step ahead predictive density of the state variable is computed as a by-product. Second, examining the filtered and smoothed volatilities for a given model provides complementary information for the specification tests that are available with EMM. EMM compares and ranks different model specifications based on an omnibus test statistic.

Although EMM is known to be efficient for many types of models, Monte Carlo studies for RS models in Gu and Zivot (2006) showed that misspecified structural models may pass the EMM specification test. This implies that ranking models using p-values alone is not enough to compare models, especially non-nested models. Estimates of the corresponding volatility can help provide more insight about the relative fit of the models. A successful model should not only provide a good fit to the data (implying a higher p-value from EMM), but also present reasonable volatilities that reflect the observed data and the underlying macroeconomic forces in the 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.

The remainder of this paper is organized as follows. Section 2 describes the models that we consider in this paper and our motivation for choosing them. A description of the particle filtering and HMM filtering algorithms needed to estimate the latent state variables is provided in Section 3. Section 4 briefly summarizes the EMM methodology and the parameter estimation results. Section 5 discusses the filtered and smoothed estimates for our models and section 6 concludes.

2 Models for the U.S. short rate

We consider two continuous time SV models and two discrete time RS models from Gu and Zivot (2006). In a study that compared a large number of volatility models, their work indicated that each of these models provided a good fit for the U.S. short rate by passing the EMM specification test.

2.1 Continuous time stochastic volatility models

We consider two continuous time diffusions

$$dr_t = (\phi_0 - \phi_1 r_t)\, dt + r_t^{\gamma} \exp\left(\frac{x_t}{2}\right) dW_{1t} = \kappa(\mu - r_t)dt + r_t^{\gamma} \exp\left(\frac{x_t}{2}\right) dW_{1t} \quad (1)$$

$$dx_t = (\omega_0 + \omega_1 x_t)dt + \xi dW_{2t} \quad (2)$$

and

$$dr_t = \phi_1 (\mu_t - r_t)\, dt + r_t^{\gamma} \exp\left(\frac{x_t}{2}\right) dW_{1t} = \kappa(\mu_t - r_t)dt + r_t^{\gamma} \exp\left(\frac{x_t}{2}\right) dW_{1t} \quad (3)$$

$$dx_t = (\omega_0 + \omega_1 x_t)dt + \xi dW_{2t} \quad (4)$$

$$d\mu_t = (\upsilon_0 + \upsilon_1 \mu_t)dt + \varsigma dW_{3t} \quad (5)$$

where r_t is the short rate at time t and dW_{1t}, dW_{2t}, and dW_{3t} are independent standard Brownian motions. The two-factor and three-factor SV models (hereafter, the SV2 and SV3 model) are extensions of the CKLS model from Chan *et. al.* (1992). The conditional mean and variance within the CKLS depend on the level of the series, which is the key characteristic of the model. Specifically, r_t mean-reverts towards the long-run level μ, with the speed of mean reversion given by κ, while γ captures the influence of the level of the series on the conditional variance, called the level effect. In the spirit of Taylor (1986, 1994), the SV models introduce additional latent factors on top of the CKLS model. The SV2 model (1) allows the log-volatility of the short rate x_t to follow a mean reverting process. The SV3 model (3) introduces another factor allowing the long-run mean to vary over time. The sensitivity of shocks to the log-volatility and to the long-run mean are measured by ξ and ς, respectively.

2.2 Regime switching models

We also consider two successful RS models

$$\Delta r_t = \phi_0 - \phi_1 r_{t-1} + \sigma_i \varepsilon_t \quad (6)$$

and

$$\Delta r_t = \phi_0 - \phi_1 r_{t-1} + \sigma_i r_{t-1}^{\gamma} \varepsilon_t \quad (7)$$

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

$$P = \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{bmatrix}$$

The entries in the matrix p_{ij} are the probabilities of transitioning from regime i to regime j.

The simple RS-in-σ model of (6) is a regime switching version of the discrete time Ornstein-Uhlenbeck (OU) process. The dynamics are assumed to be mean reverting, but regime switching is allowed for the conditional variance. Volatility shocks are regime-dependent in order to accommodate time-varying volatility. Built upon the generalized CKLS process, the RS-in-σ+*level* model of (7) incorporates the level effect as well as the RS-in-volatility effect to accommodate its 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.

3 Filtering and smoothing algorithms

This section reviews algorithms needed to estimate the latent state variables in the four models of Section 2. Section 3.4 provides details on how to implement a particle filter for the SV3 model.

3.1 General filtering and smoothing framework

State space models provide a convenient framework for analyzing the evolution of a dynamic system by formulating it into two models called the transition and measurement equations. The state variable x_t is Markovian and carries with it information about the dynamics of the system. It evolves through time via the transition equation

$$x_t = f_t\left(x_{t-1}, v_t\right) \tag{8}$$

The measurement equation then relates the state variable to the observations on the system

$$y_t = g_t\left(x_t, \varepsilon_t\right) \tag{9}$$

where v_t and ε_t are i.i.d. random variables with known probability density functions. The functions $f_t(\cdot)$ and $g_t(\cdot)$ are possibly nonlinear but known.

Associated with (8) and (9) are the transition $p(x_t|x_{t-1})$ and measurement densities $p(y_t|x_t)$ which are determined by the density of v_t and ε_t. In general, the state variable x_t can be either discrete-valued, continuous-valued or a mixture of the two. All four models in our paper require (9) to be expanded to

$$y_t = g_t(x_t, y_{t-1}, \varepsilon_t) \tag{10}$$

which technically takes them outside of the class of hidden Markov models.[2] However, this does not change any of the algorithms in this paper. For simplicity, we assume that the states are continuous and that (9) will suffice; see, Cappé *et. al.* (2005) for a general framework that covers all settings at once.

All relevant information about the unobservable or hidden states $\{x_j\}_{j=0}^T$ given observations up to and including T, $\{y_j\}_{j=1}^T$, can be obtained from the joint posterior distribution $p(x_{0:T}|y_{1:T})$ of the state variables. Estimating recursively in time three of its marginal densities is the main goal of the forecasting, optimal filtering, and signal extraction literature. The first marginal is the one-step ahead predictive density of the state

$$p(x_t|y_{1:t-1}) = p(x_t|y_1, ..., y_{t-1})$$

which uses information up until time $t-1$ to make a one period forecast of the state. The second marginal is the filtering distribution

$$p(x_t|y_{1:t}) = p(x_t|y_1, ..., y_t)$$

which carries contemporaneous information. Finally, the smoothing distribution uses all the information in the sample to estimate past realizations of the state

$$p(x_t|y_{1:T}) = p(x_t|y_1, ..., y_T)$$

[2] A hidden Markov model requires that y_t is independent conditional on x_t which is not satisfied by (10).

Obtaining the filtering density consists of two steps: prediction and update. The prediction step computes the one-step ahead predictive density of the state variable at time t based on time $t-1$ information using the Chapman-Kolmogorov equation

$$\begin{aligned} p\left(x_t \mid y_{1:t-1}\right) &= \int p\left(x_t \mid x_{t-1}, y_{1:t-1}\right) p\left(x_{t-1} \mid y_{1:t-1}\right) \mathrm{d}x_{t-1} \\ &= \int p\left(x_t \mid x_{t-1}\right) p\left(x_{t-1} \mid y_{1:t-1}\right) \mathrm{d}x_{t-1} \end{aligned} \tag{11}$$

where $p\left(x_t \mid x_{t-1}\right)$ is the transition density from above. When the measurement y_t becomes available at time t, the update step combines (11) with the full information (or data-augmented) likelihood to compute the posterior density for the current state via Bayes' rule

$$p\left(x_t | y_{1:t}\right) = \frac{p\left(y_t \mid x_t\right) p\left(x_t \mid y_{1:t-1}\right)}{p\left(y_t \mid y_{1:t-1}\right)} \tag{12}$$

where the denominator of (12) is the normalizing constant

$$p\left(y_t \mid y_{1:t-1}\right) = \int p\left(y_t \mid x_t\right) p\left(x_t \mid y_{1:t-1}\right) \mathrm{d}x_t \tag{13}$$

Optimal filters, i.e. algorithms that are the exact solutions to the recursions (11) and (12), are analytically available in only three known cases. One is the Kalman filter, which computes the marginal distributions exactly when the functions $f_t\left(\cdot\right)$ and $g_t\left(\cdot\right)$ are linear and both v_t and ε_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 state hidden Markov model (HMM) and the filter is the HMM filter, which we review in section 3.5 below.

The source of the problem in nonlinear, non-Gaussian settings is the insoluble integral in (13) that is needed to compute the one-step ahead predictive density of the observed series. The Kalman filter computes this density recursively to form the likelihood function via the prediction error decomposition; see, e.g. Durbin and Koopman (2001). This is also the density that Gallant and Tauchens' (1998) reprojection technique recovers after estimating the parameters of the structural model. In other words, their reprojection step works backwards and computes a Monte Carlo estimate of $p\left(y_t \mid y_{1:t-1}\right)$; the density that would have resulted in their parameter estimates.

3.2 Particle filters

The assumptions needed above to compute an optimal filter may not hold in situations of practical interest; for example, many dynamic systems are better modeled as nonlinear and/or non-Gaussian. Therefore, enormous efforts have been devoted to approximating the filtering and smoothing distributions when it is impossible to evaluate them analytically. We focus on particle filters and smoothers in this paper, which have shown to be reliable; see, e.g. Doucet *et. al.* (2001), Cappé *et. al.* (2005), and Cappé *et. al.* (2007) for more detailed reviews. Robert and Casella (2004) provides a general introduction to importance sampling and sequential importance sampling.

3.2.1 Particle filtering via sequential importance sampling with resampling

Particle filters are based on a Monte Carlo simulation method known as importance sampling. They use importance sampling to approximate the three marginal distributions sequentially using a set of weighted samples. Computation of the weights and drawing of the random samples can be illustrated as follows. Suppose there exists a probability density of interest called the target density $\pi(x)$ from which it is impossible to directly draw samples but that can be evaluated up to a constant of proportionality. The main idea behind importance sampling is to draw random samples 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 random draws from a proposal density $q(x)$, called the *importance density*, the realized sample $\{x^i\}_{i=1}^N$ gets reweighted by calculating importance weights

$$w^i \propto \frac{\pi(x^i)}{q(x^i)}$$

for each draw. After normalizing the importance weights $\sum_{i=1}^N w^i = 1$, the approximation to the density is given by

$$\pi(x) \approx \sum_{i=1}^{N} w^i \delta\left(x - x^i\right)$$

where $\delta(\cdot)$ denotes the Dirac mass. This weighted approximation of $\pi(x)$ converges to the true density as $N \to \infty$.

Therefore, the collection of points $\{x_t^i, w_t^i\}_{i=1}^N$ called *particles* characterizes the marginal density

$p\left(x_t|y_{1:t}\right)$, where $\{x_t^i\}_{i=1}^N$ represent points on the support of the density and the associated weights $\{w_t^i\}_{i=1}^N$ are the probability masses at those points. At each iteration, the posterior density at t will be approximated as

$$p\left(x_t|y_{1:t}\right) \approx \sum_{i=1}^{N} w_t^i \delta\left(x_t - x_t^i\right)$$

With the addition of a new observation, the next posterior density can be approximated by taking the existing particles that were simulated in the past and moving them to a new position on the support of the next density. Using previously simulated particles, transforms a basic importance sampling algorithm into a sequential importance sampling (SIS) algorithm. 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 $q\left(x_t|y_{1:t}, x_{t-1}^i\right)$, then the importance weights at any point in time are

$$w^i \propto \frac{p\left(x_t^i|y_{1:t}\right)}{q\left(x_t^i|y_{1:t}, x_{t-1}^i\right)} \tag{14}$$

which is the ratio of the target density over the importance density. In practice, the importance weights can be updated recursively to avoid calculating the numerator at each iteration. The numerator in (14) can be rewritten as

$$\begin{aligned} p\left(x_t^i|y_{1:t}\right) &= \frac{p\left(y_t \mid x_t^i\right) p\left(x_t^i \mid y_{1:t-1}\right)}{p\left(y_t \mid y_{1:t-1}\right)} \\ &= \frac{p\left(y_t \mid x_t^i\right) p\left(x_t^i \mid x_{t-1}^i, y_{1:t-1}\right) p\left(x_{t-1}^i \mid y_{1:t-1}\right)}{p\left(y_t \mid y_{1:t-1}\right)} \\ &\propto p\left(y_t \mid x_t^i\right) p\left(x_t^i \mid x_{t-1}^i\right) p\left(x_{t-1}^i \mid y_{1:t-1}\right) \end{aligned} \tag{15}$$

Substituting (15) into (14), the equation for updating the importance weights recursively is

$$\begin{aligned} w^i &\propto \frac{p\left(y_t \mid x_t^i\right) p\left(x_t^i \mid x_{t-1}^i\right) p\left(x_{t-1}^i \mid y_{1:t-1}\right)}{q\left(x_t^i|y_t, x_{t-1}^i\right) q\left(x_{t-1}^i|y_{1:t-1}\right)} \\ &= w_{t-1}^i \frac{p\left(y_t \mid x_t^i\right) p\left(x_t^i \mid x_{t-1}^i\right)}{q\left(x_t^i|y_t, x_{t-1}^i\right)} \end{aligned} \tag{16}$$

Unfortunately, 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 is known as degeneracy in the literature. In order to solve this problem, Gordon *et. al.* (1993) introduced a resampling step to the SIS algorithm which replicates better particles and eliminates particles with small probability mass. The addition of the resampling step to the SIS algorithm creates the SISR algorithm, provided as algorithm 1 in the appendix.

The resampling step copies "better" particles at each iteration in order to explore the support of the next posterior density. The simplest resampling algorithm is multinomial resampling, which corresponds to drawing new particles from a multinomial distribution with probabilities equal to the normalized importance weights. Improvements on this method have been made and are reviewed in Cappé *et. al.* (2005). It is not optimal to resample at each iteration as resampling increases the amount of Monte Carlo variation introduced into the estimates. Instead, it should be conducted only when the variance of the importance weights grows. Liu and Chen (1998) introduced a measure called the effective sample size (ESS), defined as

$$ESS = \frac{1}{\sum_{i=1}^{N} \left(w_t^i\right)^2} \tag{17}$$

to randomly determine when to resample. Criterion other than the ESS can also be used, see Cappé *et. al.* (2005) for details.

The most important part of any particle filtering algorithm is the choice of the importance density. Ideally, the researcher should choose an importance density that approximates the target density as closely as possible. This will keep the variance of the importance weights low. A convenient choice used by many authors is the transition density of the state variables which is the prior density in a Bayesian framework

$$q\left(x_t|x_{t-1}^i, y_{1:t}\right) = p\left(x_t|x_{t-1}^i\right) \tag{18}$$

Substituting (18) into (16) leads to a particularly simple update of the importance weights

$$w_t^i \propto w_{t-1}^i p\left(y_t|x_t^i\right) \tag{19}$$

If the particles are resampled at every iteration, the weight update simplifies further to $w_t^i \propto$

$p\left(y_t|x_t^i\right)$. This is the original particle filtering algorithm of Gordon *et. al.* (1993) known as the bootstrap filter.

Although the transition density of the state variables is convenient, it does not incorporate the current observation into the importance density. The particle filtering literature includes the concept of a conditionally optimal importance density, $q\left(x_t|x_{t-1}^i, y_t\right)$, which will minimize the variance of the importance weights. This density can only be computed analytically in special cases. Many sub-optimal approximations to it using local linearization techniques have been derived, see Doucet *et. al.* (2001). Finding an importance density that uses y_t in a computationally efficient manner can be challenging. The auxiliary particle filter (APF) of Pitt and Shephard (1999) is a popular alternative. They suggested incorporating y_t by sampling from an expanded importance density $q\left(x_t, j|y_{1:t}\right)$ that admits an approximation to the conditionally optimal density as a marginal. This technique will improve a standard particle filter when the signal to noise ratio is large. The current observation will have more information about the current state than the previous state. However, the APF can degrade a particle filter if the variance of the transition density is small relative to the observation noise. For the U.S. short rate series, we found that the APF works well because the series' mean volatility is relatively low, making the observations highly informative. We discuss this further in section 3.4.

3.3 Particle smoothers

Calculation of the marginal smoothing density $p\left(x_t|y_{1:T}\right)$ is a separate challenge in nonlinear, non-Gaussian state space models. Bayesian estimation of SV models via MCMC automatically delivers smoothed estimates of volatility as a by-product. This is not the case for EMM. The optimal filtering literature has demonstrated that the smoothing density can be calculated recursively based upon two separate principles: two-filter formula smoothing and backwards Markovian smoothing; see, e.g. Cappé *et. al.* (2005) for theoretical derivations. Both methods were developed in the 1960-1970's for the linear Gaussian state space model and their counterparts exist for particle methods.

Kitagawa (1996) proposed the first particle smoothing algorithm based upon two-filter formula smoothing. The method requires continually resampling past particles at each iteration which over time reduces the number of distinct particles that can represent the marginal smoothing density. Kitagawa and Sato (2001) suggesting truncating the lag length on the resampling step to make it operational. Consequently, this smoother is only valid for fixed lag smoothing.

Doucet *et. al.* (2000) developed a backwards Markovian particle smoother that is probabilistically equivalent to the Kalman smoothing approach most common in economics. We apply it in this paper as it is particularly simple to implement, see algorithm 2 in the appendix. This algorithm runs a particle filter forward storing the particles' locations and weights. Then, it runs backward in time calculating each marginal. The algorithm can be computationally time intensive because it is an $O\left(N^2\right)$ operation. Recent research in computing methods has demonstrated how to lower this time dramatically; see, Klass *et. al.* (2006) who apply N-body methods to particle smoothers.

Another drawback of this smoother concerns the fact that particle locations are simulated only on the forward pass, which fixes the support of the marginal smoothing density. Briers *et. al.* (2004) and Klass *et. al.* (2006) have developed a new two-filter formula particle smoother, also an $O\left(N^2\right)$ operation, that simulates new particle locations on a backward filtering pass. This smoother unfortunately cannot be applied to the SV2 and SV3 models due to the lagged value of y_{t-1} in the measurement equation when the SDE is discretized. Finally, another attractive approach for calculating the marginal smoothing density is provided by Godsill *et. al.* (2004), who invented a nonlinear, non-Gaussian equivalent of a simulation smoother. A simulation smoother takes random draws from the sequence of marginal smoothing densities after having run a filter forward. Averaging over many series of draws then provides a Monte Carlo estimate of the smoothing distributions.

3.4 Implementation for the SV3 model

In this section, we describe how to implement the SISR particle filter for the SV3 model. Choosing an importance density is the first step in implementing a particle filter. We experimented with several different importance densities for this model within the SISR and APF algorithms. We concluded that the APF was reliable, simple to implement, and computationally efficient.

For the first step in the APF, we set First, we need to know the transition densities $p\left(x_t|x_{t-1}\right)$ associated with the transition equations (4) and (5). The transition density of an OU process driven by Brownian motion is a normal distribution and can be simulated exactly without discretization error.[3] Given values from the previous time step for x_t and μ_t, the next values are simulated forward as

[3]Glasserman (2004) is a reference for these results.

$$
\begin{aligned}
x_{t+1} &= \exp\left(\frac{\omega_1 dt}{ndt}\right) x_t - \frac{\omega_0}{\omega_1}\left(1-\exp\left(\frac{\omega_1 dt}{ndt}\right)\right) + \sqrt{\frac{\xi}{-2\omega_1}\left(1-\exp\left(\frac{2\omega_1 dt}{ndt}\right)\right)} Z_{1,t+1} \\
\mu_{t+1} &= \exp\left(\frac{\upsilon_1 dt}{ndt}\right) \mu_t - \frac{\upsilon_0}{\upsilon_1}\left(1-\exp\left(\frac{\upsilon_1 dt}{ndt}\right)\right) + \sqrt{\frac{\varsigma}{-2\upsilon_1}\left(1-\exp\left(\frac{2\upsilon_1 dt}{ndt}\right)\right)} Z_{2,t+1}
\end{aligned}
$$

with $Z_{i,t+1} \sim N(0,1)$. The value of dt is determined by the frequency of the data which is $1/52$ for our weekly data set. The parameter ndt is the number of additional time steps that are simulated between each observation. In our work, we simulate an additional 25 time steps.

With this choice of the importance density, the importance weights reduce to computing the conditional likelihood $p(y_t|x_t)$. The likelihood for this model conditional on the simulated values of x_t and μ_t reduces to the transition density of (3) in a model with no latent variables. The transition density of the CKLS model is unfortunately unknown. A possible solution is to approximate the density with a normal distribution after discretizing the SDE with an Euler or Milstein scheme. The density is then evaluated as

$$
p(y_t|x_t) = N\left(y_{t-1} + \kappa\{\mu_t - y_{t-1}\}dt, \exp(x_t)\, y_{t-1}^{2\gamma} dt\right)
$$

Alternatively, the estimated value of γ in the SV3 model is near 0.5 making it close to a CIR process. The transition density of the CIR process is known to be a noncentral chi-square distribution.

We also implemented particle filters with alternative importance densities. In our simulations, we found that the APF performed significantly better than a SISR algorithm using the transition density as a proposal and moderately better than a SISR algorithm with the Kalman filter approximation of Harvey *et. al.* (1994) as proposal. We also experimented with the approximation to the conditionally optimal proposal described in Cappé *et. al.* (2005). The latter may be slightly more efficient for a fixed number of particles but not necessarily for a fixed computational time because it requires calling a mode finding algorithm for each particle at each iteration. Another one of our goals is to describe methods that a user of EMM can code reliably. More elaborate particle filters can easily be designed. We note two in particular that are relevant here. Markov chain Monte Carlo steps can be included as in Gilks and Berzuini (2001). And, Doucet *et. al.* (2006) have recently described a block sampling approach, which may provide significant efficiency gains.

There exists other work on using particle filters to estimate continuous time stochastic volatil-

ity models or continuous time models in general. Johannes *et. al.* (2006) have recently used an algorithm similar to ours to estimate stochastic volatility models with jumps. Their algorithm is essentially the bootstrap filter with a proposal density based on an Euler discretization of the transition equation. In addition, Johannes *et. al.* (2006) simulate artificial observations between observed data that they then use to compute additional importance weights. Fearnhead *et. al.* (2006) proposed an original particle filtering algorithm for partially observed diffusions based on recent advances in the simulation of diffusions. They use results from Beskos *et. al.* (2005) and Beskos *et. al.* (2006) who have developed methods for simulating specific classes of diffusions without discretization error. Finally, Rimmer *et. al.* (2005) use particle filters to compute ML estimates for the parameters of a diffusion measured in noise. Their article offers additional advice on building importance densities. The relative performance of these continuous time particle filtering algorithms has yet to be determined and is an important area of future research.

3.5 The Hidden Markov Model Filter and Smoother

In section 4 and 5, we estimate the RS models (6) and (7) that have state variables taking on discrete values and evolve over time according to a first-order Markov transition matrix. In this case, the optimal filter and smoother can be computed analytically as in Baum and Petrie (1966). Technically, our models are RS models and not HMMs because the current observation depends upon past observations as in (10). This means that conditional on the state variable the current observation is not independent like a true HMM. Hamilton's (1989) algorithm is slightly different than Baum and Petries' (1966) because his model had lagged state variables in the measurement equation. As our models do not have lagged state variables in the measurement equation, the filter and smoother can still be appropriately computed using the HMM filter and smoother.

The recursion for the HMM filter follows a two-step procedure with prediction

$$\begin{aligned} p\left(x_t|y_{1:t-1}\right) &= \sum_{j=1}^{K} p\left(x_t = j|y_{1:t-1}\right) \\ p\left(x_t = j|y_{1:t-1}\right) &= \sum_{i=1}^{K} p\left(x_{t-1} = i|y_{1:t-1}\right) p_{ij} \end{aligned}$$

and update

$$
\begin{aligned}
p\left(x_t|y_{1:t}\right) &= \sum_{j=1}^{K} p\left(x_t = j|y_{1:t}\right) \\
p\left(x_t = j|y_{1:t}\right) &= \frac{p\left(x_t = j|y_{1:t-1}\right) p\left(y_t|x_t = j\right)}{\sum_{k=1}^{K} p\left(x_t = k|y_{1:t-1}\right) p\left(y_t|x_t = k\right)} \qquad (20)
\end{aligned}
$$

Here, there are K possible states and as above p_{ij} is the probability of transitioning from state i to state j.[4] The HMM smoothing algorithm based upon the two-filter formula smoothing principle follows as

$$
\begin{aligned}
p\left(x_{t|T} = i|y_{t:T}\right) &= \frac{\sum_{j=1}^{K} q_{ij} p\left(y_{t+1}|x_{t+1} = j\right) p\left(x_{t+1|T} = j|y_{t:T}\right)}{\sum_{k=1}^{K} p\left(x_t = k|y_{1:t-1}\right) p\left(y_t|x_t = k\right)} \\
p\left(x_t = j|y_{1:T}\right) &= \frac{p\left(x_t = j|y_{1:t}\right) p\left(x_{t|T} = j|y_{t:T}\right)}{\sum_{k=1}^{K} p\left(x_t = k|y_{1:t}\right) p\left(x_{t|T} = k|y_{t:T}\right)}
\end{aligned}
$$

3.5.1 Viterbi Algorithm

We also implement the Viterbi (1967) algorithm for this set of models, given as algorithm 3 in the appendix. This is a dynamic programming algorithm that computes the maximum a posteriori (MAP) estimator of the sequence of state variables. That is, it computes the *joint sequence* of states $x_{1:T}$ that maximizes the conditional probability of the data $p\left(x_{1:T}|y_{1:T}\right)$. This algorithm answers a fundamentally different question than a smoothing algorithm which computes the marginal distribution of an individual state $p\left(x_t|y_{1:T}\right)$. The MAP estimator is rarely used in the economics literature, even when its output is easier to interpret. The Viterbi algorithm is provided as algorithm 3.

Although it was originally derived for discrete state models, it can also be calculated for the SV2 and SV3 models by particle methods; see, e.g. Godsill *et. al.* (2001) and Klass *et. al.* (2006). We apply this method for the sake of completeness. Note that the conditional mean of the marginal $p\left(x_t|y_{1:T}\right)$ remains the minimum MSE estimator, while the MAP estimator may be preferable and noticably different in settings of multi-modality where the mean exists in a lower probability region between the modes.

[4] Hamilton's algorithm includes an additional sum in the denominator of (20) to account for lagged state variables.

4 EMM estimation results

In this section, the EMM methodology is briefly described and estimation results for fitting the U.S. short rate are summarized.

4.1 EMM methodology

The theory of EMM was developed in Gallant and Tauchen (1996) and was extended to non-Markovian data with latent variables in Gallant and Long (1997). Gallant and Tauchen (2001) and Zivot and Wang (2006) provide overviews of the methodology in more detail. The basic procedure of EMM estimation consists of two steps. First, the empirical conditional density of the 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 objective function is constructed using the score functions (from the log-likelihood of the SNP density) as moments. The scores are evaluated using data simulated from a given structural model and the objective function is minimized with respect to the parameters underlying the structural model. This step extracts structural parameters from the data by minimizing the χ^2 criterion and is therefore called the estimation step.

4.2 Data

Our empirical work uses weekly observations of the annualized yield on the 3-month U.S. T-bill over the period January 1954 to September 2004, forming 2468 observations. The data was constructed from a daily series available from the U.S. Federal Reserve, where the rates are calculated as unweighted averages of closing bid rates quoted by at least five dealers in the secondary market. The rates are posted on a bank discount basis, but converted to continuously compounded yields prior to analysis. Wednesday data are used because it results in the smallest number of missing observations. The raw data are plotted in Figure 1 and descriptive statistics are given in Table 1. The data set is similar to Gallant and Tauchens' (1998) and has been extended for five additional years; a time period that featured reduced volatility and a lower level. The basic stylized facts are: near non-stationary behavior (slow mean reversion), large changes and small changes are clustered

together (volatility clustering), the volatility increases with the level (level effect), and positive skewness and excess kurtosis (non-Gaussian distribution).

4.3 EMM estimation results for the SV models

The preferred SNP model for describing the short rate data is an AR(1)-GARCH(1,1)-$Kz(7)$ model and the EMM estimation results for the two SV models are reported in Table 2. This is slightly different than the SNP model fit by Gallant and Tauchen (1998), who use an AR(1)-ARCH(4)-$Kz(?)-Kx(?)$. Gu and Zivot (2006) report that the CKLS model without additional latent factors gets rejected easily. On the other hand, the SV2 model and the SV3 model were not rejected at the 10% level, implying that the introduction of additional stochastic factor(s) is important for explaining the behavior of the U.S. short rate. The SV3 model was even favored over the SV2 model; the fit improved further when adding a stochastic mean.

Strong mean reverting behavior was confirmed for both models with similar reverting trend, measured by ϕ_0/ϕ_1 or υ_0/υ_1, around 2.5%. This estimate is lower than in previously published work, partially due to the incorporation of a 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. 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. Adjusted t-ratios of all individual elements of the score vector are well below 2.0 for the accepted SV2 and SV3 model, implying that both models have no difficulties in capturing the volatility clustering that exists in the data.

4.4 EMM estimation results for the regime-switching models

The EMM estimation results for the two RS models are provided in Table 2. 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-in-σ model by 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.

Both RS models exhibit strong mean reversion with the long-run reverting means estimated higher than those of the SV models. A low-volatility and a high-volatility regime are identified with

highly persistent transition probabilities; both probabilities exceed 0.90. The transition probability of staying in the high-volatility regime is estimated to be smaller than in previous work, e.g. Smith (2002). Lastly, the estimated conditional volatility in the RS-in-σ+*level* model is sensitive to the level of the short rate series. The magnitude of the level effect is much lower than unity but very similar to what is found in the multi-factor SV models.

5 Filtering and smoothing results

5.1 Results for the SV models

Figures 2 and 3 display the one-step ahead predicted, filtered, smoothed, and MAP estimates of the volatility from the SV2 model using the APF algorithm described in section 3.4. We used $N = 2000$ particles and the systematic resampling algorithm of Carpenter *et. al.* (1999). For the SV2 model, the filtered and smoothed volatility follow similar paths. Moreover, a number of high-volatility periods coincide with the historical business cycles and important events from 1954 to 2004. Specifically, the volatility process has two spikes in the 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. The short rate experienced another period of high volatility during the 1970's when stagflation hit the U.S. economy. Critical macroeconomic events include 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 appears relatively stable since the mid-1980's. Two significant spikes in the volatility are due to the stock market crash of October 19th, 1987 and the terrorist attack on September 11th, 2001.

Figure 4 provides the predicted, filtered, and smoothed estimates of the volatility and time-varying mean for the SV3 model. Although the SV3 model fits the data better than the SV2 model based on the results from EMM, the filtered and smoothed volatilities from both models share similar paths. The estimated volatilities from the SV3 model during the high-volatility periods are slightly higher than those from the SV2 model; for example, during the periods of the monetary experiment and the market crash of October 19th, 1987.

The filtered estimates of the time-varying mean μ_t in the SV3 model are heavily revised when

compared to the smoothed estimates. Specifically, the filtered estimates are consistently lower than the smoothed estimates. For example, the U.S. economy experienced an unprecented period of economic growth in the late 1990's. During this time, the volatilities of the short rate were enduringly low but the reverting means kept increasing. When this period of growth was ended by the terrorist attack on Sept. 11th, 2001, the short rate became more volatile while its mean dropped significantly as the Fed reacted to world events.

5.2 Results for the regime switching models

We apply the HMM filter and smoother and the Viterbi algorithm to our regime switching models (6) and (7). The results are shown in Figures 6 and 7 for the RS-in-σ and RS-in-σ+*level* models, respectively. The panels display the filtered and smoothed probabilities from the HMM algorithm and finally the sequence of joint states from the Viterbi algorithm. Although the RS-in-σ+*level* model fits the data better than the RS-in-σ model based on the p-value from EMM, our filtering and smoothing results provide evidence that the RS-in-σ model better reflects the business cycles and important events that may cause strong shifts in the behavior of interest rates.

For both the RS-in-σ+ and RS-in-σ+*level* models, the estimated switching probability is noisy using the information up to the current state, but improves greatly when all the data gets incorporated; see the smoothed probabilities and the joint sequence of states. The MAP estimator implies that there are twelve high-volatility regimes from 1954-2004. Specifically, the second and third high-volatility regimes correspond to the two short recessions during 8/1957 - 5/1958 and 4/1960 - 2/1961; the following three high-volatility regimes appear in the early 1970s when recession and inflation ocurred simultaneously, 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 U.S. Federal Reserve during 1979-82. Afterwards, the U.S. economy experienced a long low-volatility period in the 1990's with strong growth. Our estimation successfully captures two abnormal events that interrupted it: one is the stock market crash of October 19th, 1987; the other is the terrorist attack on Sept. 11th, 2001.

Figure 7 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. From this model, more high volatility regimes have been identified and the estimated indicator switches between regimes much more frequently than 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 business cycle dates and other important historical events. This is the case even when the RS-in-σ+*level* model provided the highest p-value of all the volatility models we considered. Specifically, the RS-in-σ+*level* model indicates that the interest rate stays in the high-volatility regime 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 in the 1950's and early 1960's, are not clearly identified. Moreover, the two high-volatility regimes, can be easily observed from the time series of the change in the data. Several additional high-volatility regimes are also identified, although they only lasted short periods of time. The well-known period of the monetary experiment during 1979-1982 appears to expand to 1985 according to the estimated regimes obtained from this model. Additional high-volatility regimes are also implied during the 1990's, although the U.S. economy and the short rate are relatively stable during that period. Based on all the above evidence, our estimation results show that the RS-in-σ+*level* model over-fits the historical data.

6 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 sizeable literature that uses EMM to estimate the parameters of volatility models, estimating the volatility has not been widely explored within this literature. In this paper, we evaluate the performance of a variety of volatility models in terms of the filtered and smoothed estimates of the latent processes using appropriate algorithms from the optimal filtering and signal extraction literature. For the continuous time SV models, the estimated volatilities follow similar dynamic paths. In addition, the filtered and smoothed volatilities from either model can resemble the historical business cycles and important events during 1954 to 2004. 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 cycle and important events than from the RS-in-σ model. The former model provided the best fitting performance across all the models considered during EMM estimation. Our results indicate an over-fit of the time series for the RS-in-σ+*level*

model.

References

[1] Anderson, T. G., Chung, H. J., and Sorensen, B. E. (1999). Efficient method of moments estimation of a stochastic volatility model: a Monte Carlo study. *Journal of Econometrics*, Vol. 91, pp. 61-87.

[2] Anderson, T. G., and Bollerslev, T. (2005). A framework for exploring the macroeconomic determinants of systematic risk. Manuscript, Annual Meeting of the American Economic Association.

[3] Baum, L. E., and Petrie, T. P. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics*, Vol. 37, pp. 1554-1563.

[4] Baum, L. E., Petrie, T. P., Soules, G., and Weiss, N. (1970). A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, Vol. 41, pp. 164-171.

[5] Beskos, A., and Roberts, G. O. (2005). Exact simulation of diffusions. *Annals of Applied Probability*, Vol. 15, No. 4, pp. 2422-2444.

[6] Beskos, A., Papaspiliopoulos, O., Roberts, G. O., and Fearnhead, P. (2006). Exact and efficient likelihood based inference for discretely observed diffusions. *Journal of the Royal Statistical Society*, Series B, pp. 333-382.

[7] Briers, M., Doucet, A., and Maskell, S. (2004). Smoothing algorithms for state space models. *Technical Report*, University of Cambridge, Department of Engineering.

[8] Cappé, O., Moulines, E., and Rydén, T. (2005). *Inference in Hidden Markov Models*. New York, NY: Springer-Verlag.

[9] Cappé, O., Godsill, S., and Moulines, E. (2007). An overview of existing methods and recent advances in sequential Monte Carlo. *IEEE Proceedings in Signal Processing*, forthcoming.

[10] Carpenter, J., Clifford, P., and Fearnhead, P. (1999). Improved particle filter for non-linear problems. *IEE Proceedings. Part F: Radar and Sonar Navigation*, Vol. 146, No. 1, pp. 2-7.

[11] Chernov, M., Gallant, R., Ghysels, E., and Tauchen, G. (2003). Alternative models for stock price dynamics. *Journal of Econometrics*, Vol. 116, pp. 225-257.

[12] Chib, S., Nardari, F., and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, Vol. 108, pp. 281-316.

[13] Doucet, A., Briers, M., and Sénécal, S. (2006). Efficient block sampling strategies for sequential Monte Carlo. *Journal of Computational and Graphical Statistics*, Vol. 15, no. 3, pp. 693-711.

[14] Doucet, A., de Freitas, J. F. G., and Gordon, N. J., Eds. (2001). *Sequential Monte Carlo Methods in Practice*. New York, NY: Springer-Verlag.

[15] Doucet, A., Godsill, S., and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, Vol. 10, No. 3, pp. 197-208.

[16] Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump diffusions. *Econometrica*, Vol. 68, pp. 1343-1376.

[17] Durbin, J., and Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford, UK: Oxford University Press.

[18] Fearnhead, P., Papaspiliopoulos, O., and Roberts, G. (2006). Particle filters for partially observed diffusions. Manuscript, Department of Mathematics, Lancaster University.

[19] Gallant, A. R., Hsieh, D. A., and Tauchen, G. (1997). Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics*, Vol. 81, No. 1, pp. 159-192.

[20] Gallant, A. R., and Long, J. R. (1997). Estimating stochastic differential equations efficiently by minimum Chi-square. *Biometrika*, Vol. 84, pp. 125-142.

[21] Gallant, A. R., and Tauchen, G. (1996). Which moments to match? *Econometric Theory*, Vol. 12, pp. 657-681.

[22] Gallant, A. R., and Tauchen, G. (1998). Reprojecting partially observed systems with application to interest rate diffusions. *Journal of the American Statistical Association*, Vol. 93, pp. 10-24.

[23] Gallant, A. R., and Tauchen, G. (2001). Efficient method of moments. Manuscript, Department of Economics, Duke University.

[24] Gallant, A. R., and Tauchen, G. (2001). SNP: A program for nonparametric time series analysis, Version 8.8, User's Guide. Manuscript, Department of Economics, Duke University.

[25] Gallant, A. R., and Tauchen, G. (2002). EMM: A program for efficient method of moments estimation, Version 1.6, User's Guide. *Manuscript*, Department of Economics, Duke University.

[26] Gilks, W. R., and Berzuini, C. (2001). Following a moving target - Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society*, Series B, Vol. 63, No. 1, pp. 127-146.

[27] Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering*. New York, NY: Springer-Verlag.

[28] Godsill, S., Doucet, A., and West, M. (2001). Maximum a posteriori sequence estimation using Monte Carlo particle filters. *Annals of the Institute of Statistical Mathematics*, Vol. 53, pp. 82-96.

[29] Godsill, S., Doucet, A., and West, M. (2004). Monte Carlo smoothing for nonlinear time series. *Journal of the American Statistical Association*, Vol. 99, pp. 156-168.

[30] Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993). A novel approach to nonlinear/non-Gaussian state estimation. *IEE Proceedings - F*, Vol 140, No. 2, pp. 107-113.

[31] Gu, Y., and Zivot, E. (2006). A comparison of univariate stochastic volatility models for the U.S. short rate using EMM. Manuscript, Department of Economics, University of Washington.

[32] Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, Vol. 57, pp. 357-384.

[33] Harvey, A. C., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, Vol. 61, No. 2, pp. 247-264.

[34] Johannes, M., Polson, N., and Stroud, J. (2006). Optimal filtering of jump-diffusions: extracting latent states from asset prices. Manuscript, Columbia University, Graduate School of Business.

[35] Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies*, Vol. 65, No. 3, pp. 361-393.

[36] Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, Vol. 5, pp. 1-25.

[37] Kitagawa, G., and S. Sato. (2001). Monte Carlo smoothing and self-organizing state-space model. A. Doucet, J. F. G. de Freitas, and N. J. Gordon, Eds. *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag.

[38] Klass, M., Briers, M., de Freitas, N, Doucet, A., Maskell, S., and Lang, D. (2006). Fast particle smoothing: if I had a million particles. *Proceedings of the 23rd International Conference on Machine Learning*, Pittsburgh, PA, June 25-29, 2006.

[39] Liu, J., and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, Vol. 93, pp. 1032-1044.

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

[41] Rimmer, D., Doucet, A., and Fitzgerald, W. J. (2005). Particle filters for stochastic differential equations of nonlinear diffusions. *Technical report* # 497, Signal Processing Laboratory, Department of Engineering, University of Cambridge.

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

[43] Robert, C. P., and Casella, G. (2004). *Monte Carlo Statistical Methods.* Second Edition. New York, NY: Springer-Verlag.

[44] Smith, D. R. (2002). Markov switching and stochastic volatility diffusion models of short-term interest rates. *Journal of Business and Economic Statistics*, Vol. 20, No. 2, pp. 183-197.

[45] Taylor, S. J. (1986). *Modeling Financial Time Series.* Wiley Press: New York.

[46] Taylor, S. J. (1994). Modeling stochastic volatility. *Mathematical Finance*, Vol. 4, pp. 183-204.

[47] Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimal decoding algorithm. *IEEE Transactions in Information Theory*, Vol. 13, pp. 260-269.

[48] Zivot, E., and Wang, J. (2005). *Modeling Financial Time Series with S-PLUS.* Second Edition. New York, NY: Springer-Verlag.

Algorithms

Algorithm 1 Sequential Importance Sampling with Resampling (SISR)

1. Initialization: for $i = 1, ..., N$
• Draw x_0^i from its stationary distribution.

Iterate steps 2 and 3 for $t = 1, ..., T$.

2. Sampling: for $i = 1, ..., N$
• Draw x_t^i from $q\left(x_t \mid x_{t-1}^i,\, y_{1:t}\right)$.
• Calculate the importance weights w_t^i using (16).

3. Resampling: for $i = 1, ..., N$
• Normalize the importance weights: $\hat{w}_t^i = \frac{w_t^i}{\sum_{i=1}^N w_t^i}$.
• Calculate the effective sample size using (17).
• If $\widehat{N}_{eff} < N_{thresh}$ resample the particles using one of the resampling algorithms.
• If the particles were resampled, set their weights equal.

Algorithm 2 Forward-Backward Particle Smoother

1. Forward filtering: for $t = 1, ..., T$.
• Run a particle filtering algorithm forward and save $\left\{x_t^i,\, w_t^i\right\}_{i=1}^N$.

2. Initialization: at $t = T$.
• Set $w_{T|T}^i = w_T^i$ for $i = 1, ..., N$.

3. Backward smoothing: for $t = T - 1, ..., 1$.
• For $i = 1, ..., N$ compute smoothed weights:

$$w_{t|T}^i = w_t^i \left[\sum_{j=1}^N w_{t+1|T}^i \frac{p\left(x_{t+1}^j \mid x_t^i\right)}{\sum_{k=1}^N w_t^k p\left(x_{t+1}^j \mid x_t^k\right)} \right]$$

Algorithm 3 Viterbi Algorithm

1. Initialization: for $t = 0$ and for $i = 1, ..., K$.
• Set $\delta_0(i) = \log(\overline{p}_{ii})$ where $\overline{p}$ is the stationary distribution of the Markov transition matrix.

2. Forward recursion: $t = 1, ..., T$ and for $j = 1, ..., K$.
• Set $\delta_t(j) = \max_{i \in \{1,...,K\}} [\delta_{t-1}(i) + \log p(x_t = j | x_{t-1} = i)] + \log p(y_t | x_t = j)$.
• Set $\psi_t(j) = \max_{i \in \{1,...,K\}} [\delta_{t-1}(i) + \log p(x_t = j | x_{t-1} = i)]$.

3. Backward initialization: at $t = T$
• Set $\hat{x}_T = \max_{i \in \{1,...,K\}} [\delta_T(i)]$.

4. Backward recursion: $t = T - 1, ..., 1$.
• Set $\hat{x}_t = \psi_t(\hat{x}_{t+1})$.

Tables

Panel (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	

Panel (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.: 0.237	Skewness: -0.522	Kurtosis: 24.81	

Table 1: Descriptive statistics for the raw data (A) and the differenced series (B).

Parameter	SV2	SV3	RS-in-σ	RS-in-σ+*level*
ϕ_0	0.8428 (0.2669)	-	0.2580 (0.0336)	0.1769 (0.5818)
-ϕ_1	-0.2956 (0.1816)	-1.0947 (0.8463)	-0.0284 (0.1640)	-0.0283 (0.2976)
ϕ_0/ϕ_1	2.8512	-	9.0845	6.2070
γ	0.6659 (0.1163)	0.5167 (0.0755)	-	0.5076 (0.1266)
σ_1	-	-	0.1472 (0.0628)	0.0389 (0.0914)
σ_2	-	-	0.4613 (0.0566)	0.1400 (0.0874)
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)	-	-
ς	-	0.4715 (2.1440)	-	-
p_{11}	-	-	0.98	0.98
p_{22}	-	-	0.91	0.89
χ^2	10.35	5.55	11.01	6.53
p-value	0.1107	0.2357	0.0881	0.2916
d.o.f.	6	4	6	5

Table 2: Estimates with standard errors for the SV2, SV3, RS-in-σ, and RS-in-σ+*level* models.

Figures

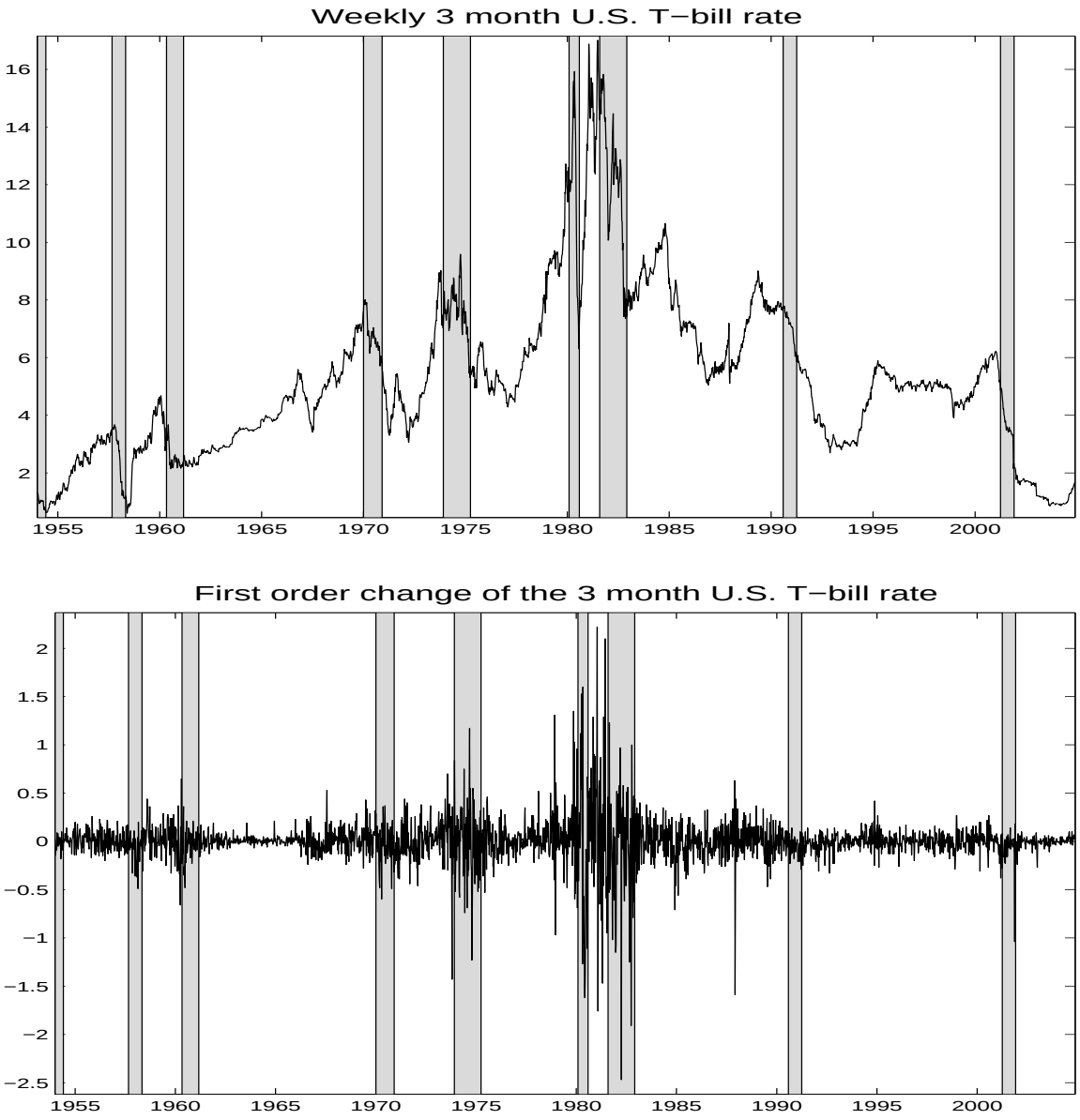

Figure 1: U.S. 3 month T-bill rate with NBER recession dates. Top: raw series. Bottom: First differenced series.

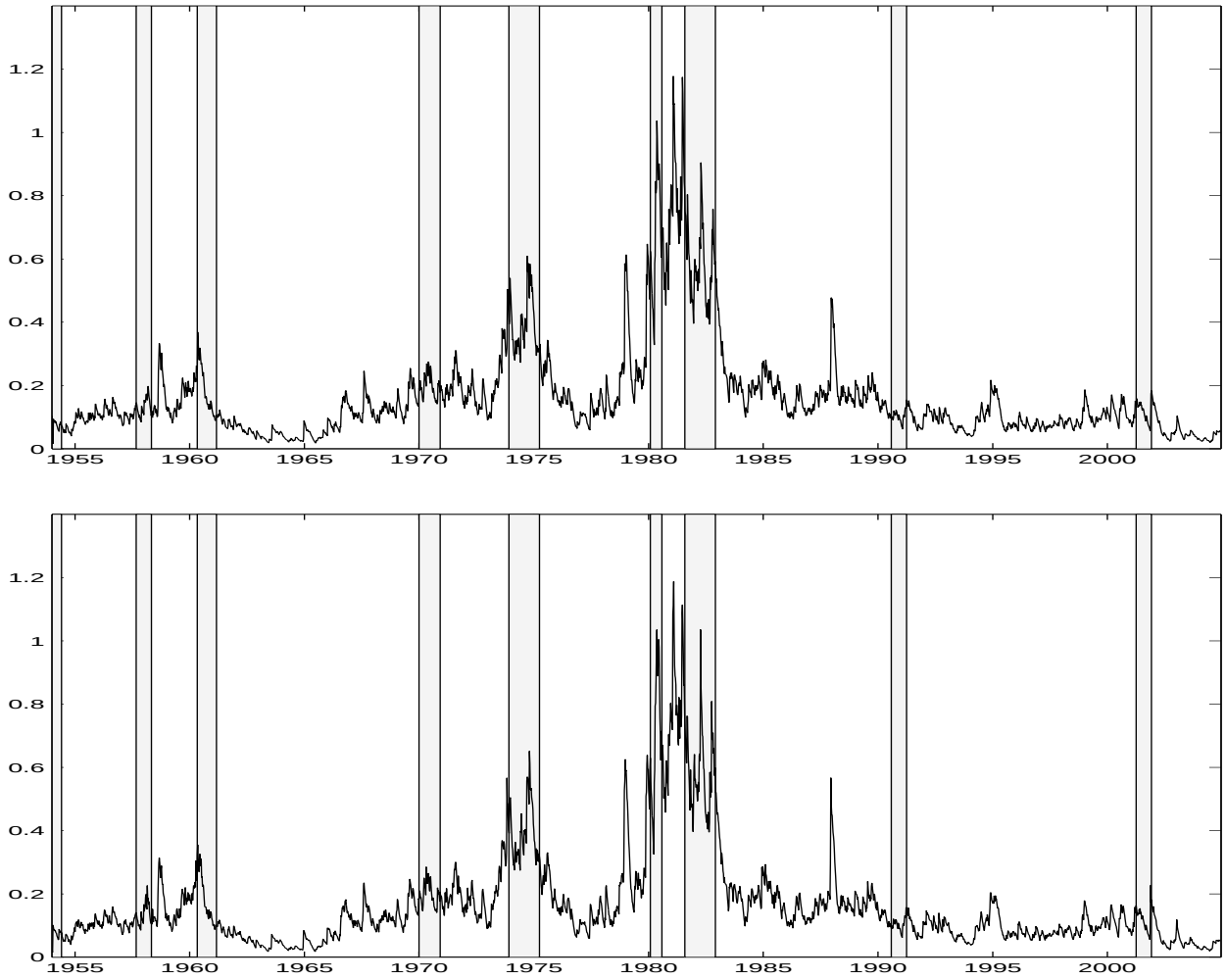

Figure 2: Estimates of conditional volatility from the SV2 model using the SISR algorithm. Top: one-step-ahead prediction. Middle: filtered.

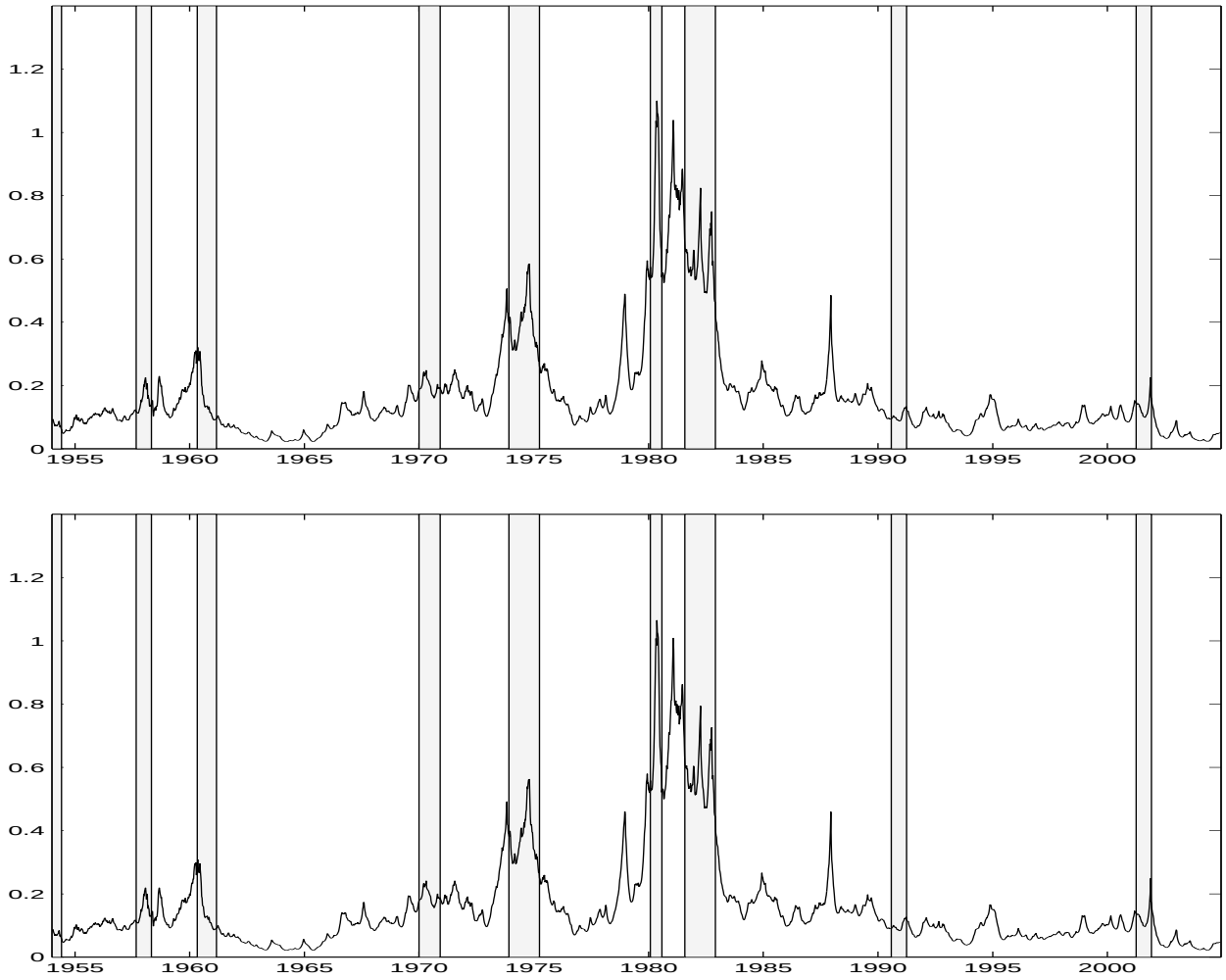

Figure 3: Smoothed and MAP estimates of conditional volatility from the SV2 model. Top: SISR smoother MAP. Bottom: SISR-MAP estimator.

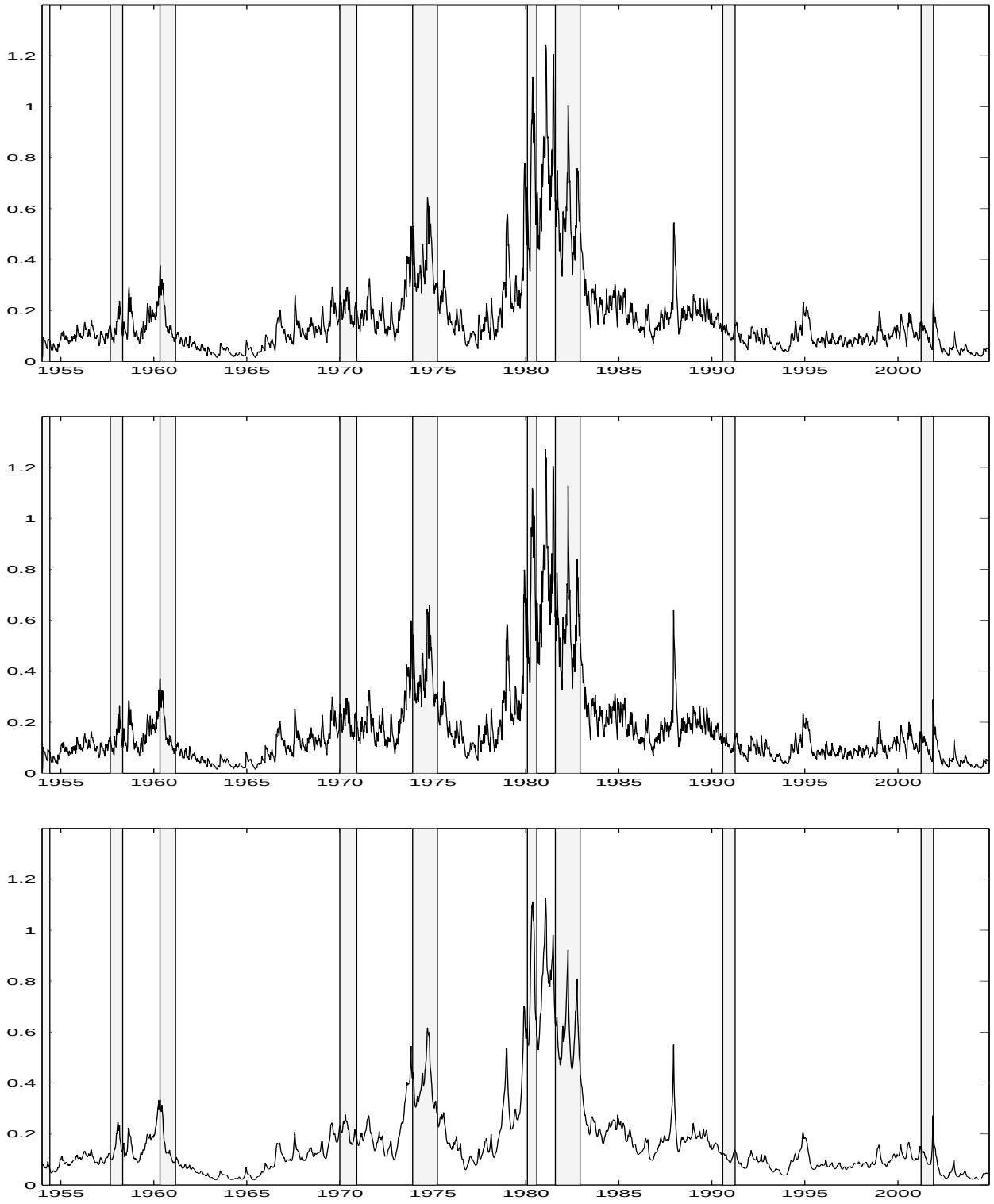

Figure 4: Estimates of volatility from the SV3 model using the SISR algorithm. Top: one-step-ahead prediction. Middle: filtered. Bottom: smoothed.

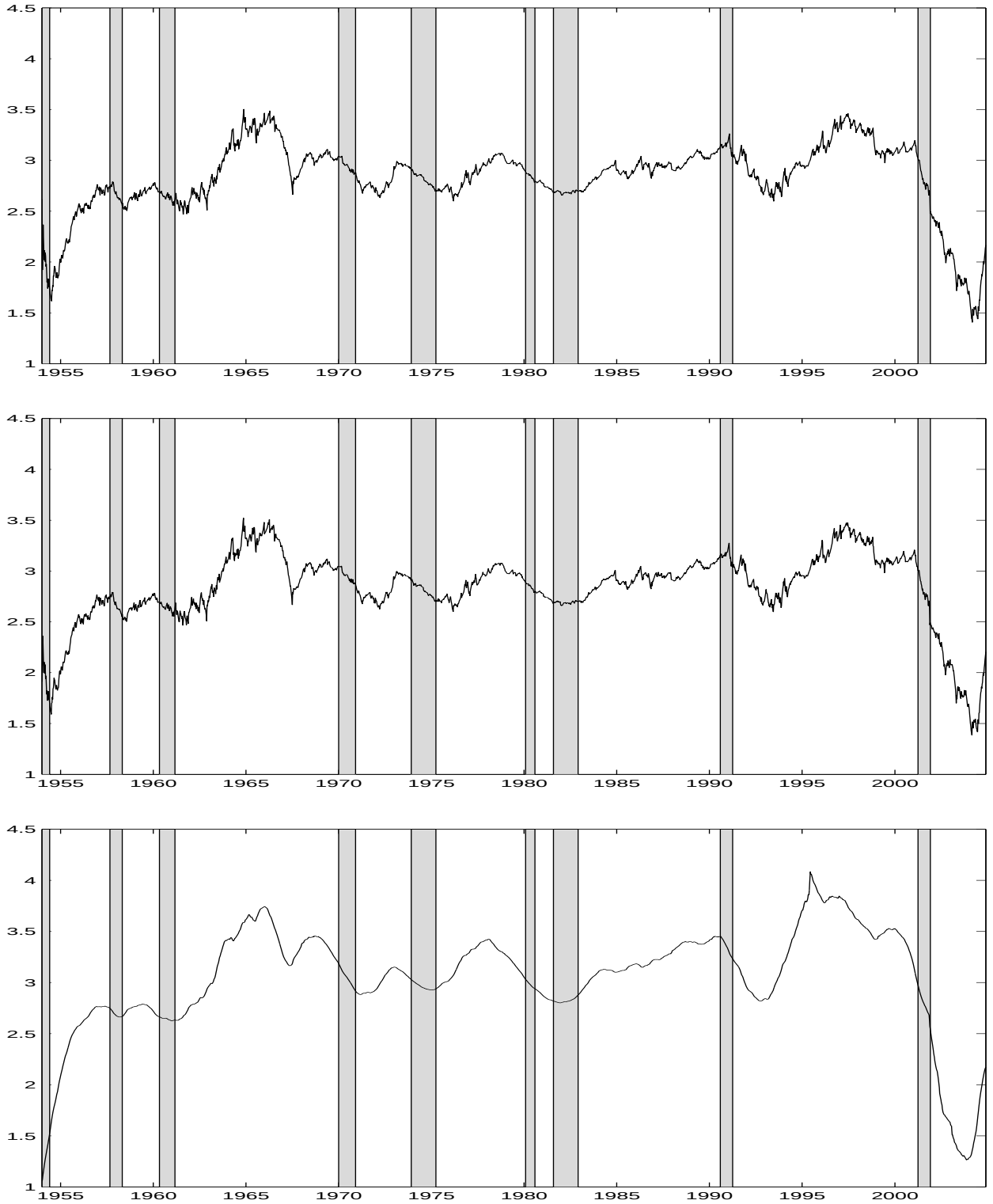

Figure 5: Estimates of the time-varying mean state variable μ_t from the SV3 model using the SISR algorithm. Top: one-step-ahead prediction. Middle: filtered. Bottom: smoothed.

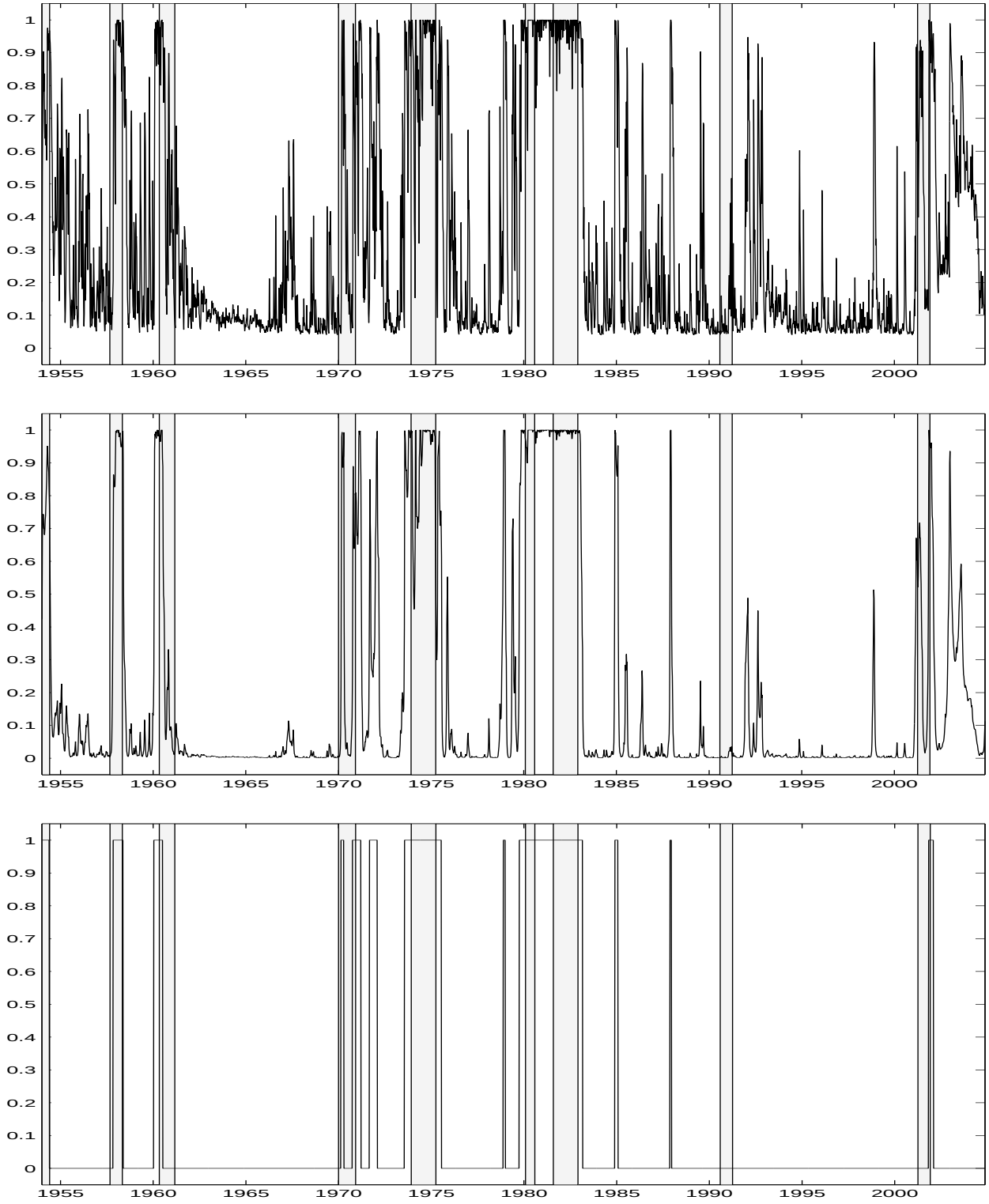

Figure 6: Estimates of the probabilities and states from the RS-in-σ model. Top: filtered probabilities. Middle: smoothed probabilities. Bottom: MAP estimate.

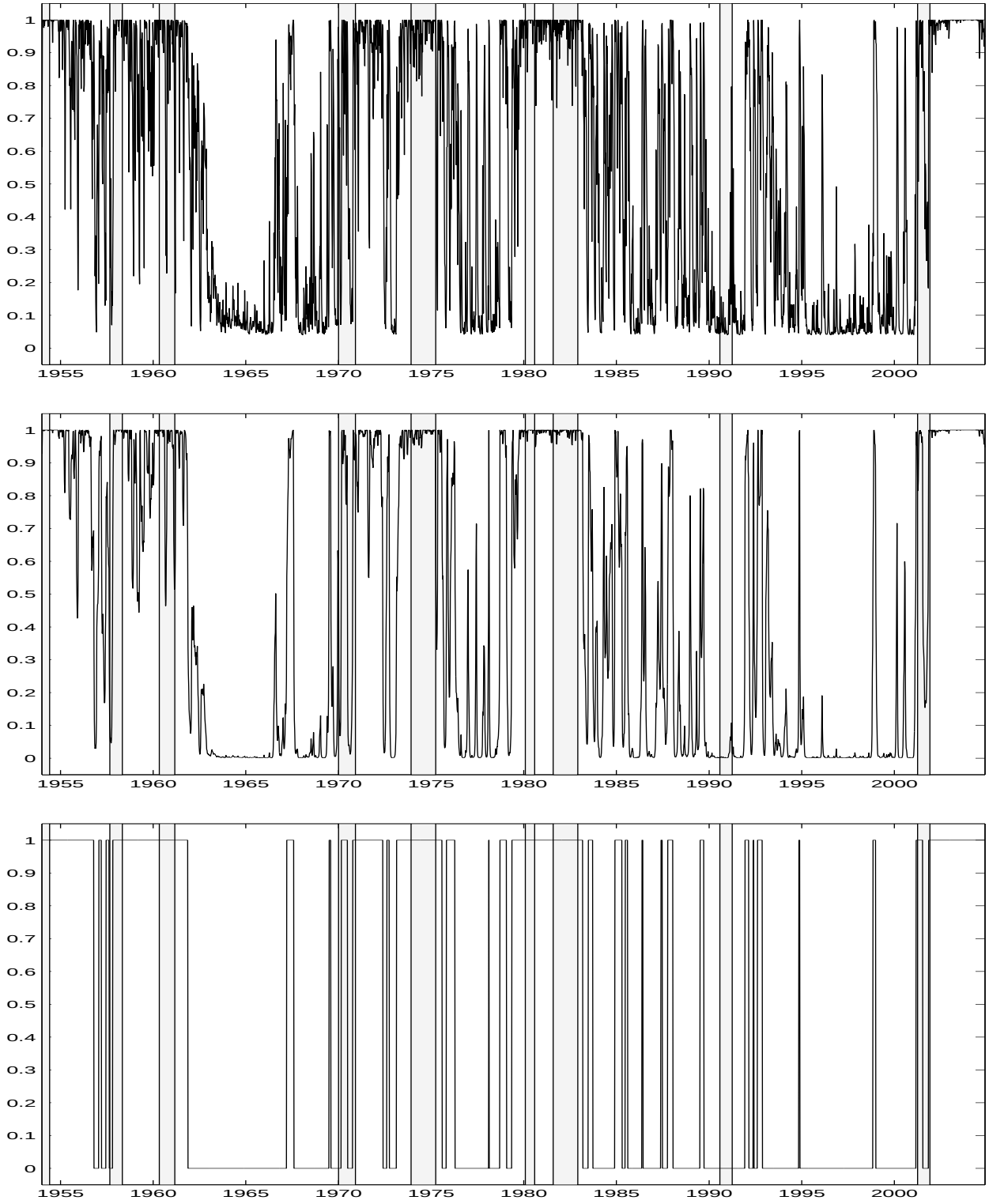

Figure 7: Estimates of the probabilities and states from the RS-in-σ+*level* model. Top: filtered probabilities. Middle: smoothed probabilities. Bottom: MAP estimate.