Markov Chain Monte Carlo Calibration of

Stochastic Volatility Models

Xin Ge *   Chuanshu Ji **

* MathSoft, Inc., 1700 Westlake Avenue North, Suite 500, Seattle, WA 98109, USA;
    `xge@email.unc.edu`

** Department of Statistics, University of North Carolina, Chapel Hill, NC 27599,
    USA; `cji@email.unc.edu`; 919-962-3917

October 10, 2000

Abstract

In this paper we propose a Bayesian computational scheme for calibration of stochastic volatility models using data from underlying asset returns and derivative prices. A special feature of this work is a combination of historical volatility and implied volatility. Markov chain Monte Carlo (MCMC) methods play a pivotal role in the simulation of historical time series for the real-world dynamics and pricing formulas driven by the risk-neutral dynamics.

JEL Classification: C11, G12

Key Words: Markov chain Monte Carlo, stochastic volatility, historical volatility, implied volatility, option pricing, risk management

1   Introduction

Quantitative finance is a fast-growing area due to the tremendous need for supporting various tasks in financial industry, such as asset pricing and risk management. The mathematical theory is developed along two parallel avenues: partial differential equations and equivalent martingale measures, with the Feynman-Kac formula serving as a bridge between them. Numerical computation and statistical inference become indispensable when implementing the theory.

Volatility calibration is at the center stage of empirical finance with a great impact on trading and pricing. Many efforts have been made to generalize the Black-Scholes theory, especially to relax the assumption of constant volatility. The volatility smile phenomenon and term structure suggest that it is more reasonable to consider volatility as a stochastic process. Two classes stand out in the research of volatility calibration: (i) ARCH/GARCH time series models [Engle (1982), Bollerslev (1986)], in which volatility is defined as a function of past returns, make the inference and forecasting easy to implement, but fail to explain certain market behaviors such as the volatility smile; (ii) Stochastic volatility (SV) models [see Taylor (1994), Shephard (1996) and references therein], in which volatility is treated as a latent process, are capable of demonstrating the volatility smile, but invite significant challenges in computation.

Frequentist methods for SV models include generalized method of moments (GMM) and efficient methods of moments (EMM) [Chesney and Scott (1989), Melino and Turn-

bull (1990), Gallant, Hsieh and Tauchen (1991), Anderson and Sorensen (1995), Gallant, Hsieh and Tauchen (1997)], quasi-maximum likelihood (QML) [Harvey, Ruiz and Shephard (1994)]. On the other hand, the Bayesian approach is a natural alternative for studying hierarchical SV models with a latent volatility layer, and its power and promise increase rapidly with the help of Markov chain Monte Carlo (MCMC) methods [Jacquier, Polson and Rossi (1994), Chen (1997), Chib, Nardari and Shephard (1998)].

A basic issue in volatility calibration concerns what data to use: underlying asset returns (e.g. stocks and bonds) or derivatives (e.g. options)? This issue mirrors two different angles from which volatility is viewed: historical volatility and implied volatility. The former infers volatility based on time series data of the underlying asset returns; the latter "inverts" a relevant pricing formula, noticeably the Black-Scholes formula for pricing European options, to obtain the (implied) volatility. Superficially, the two types of data do not appear to have essential distinctions. However, a deeper root is tied with the probability measure that generates the data. In a historical time series model, only the real-world probability measure P is relevant, while the specification of implied volatility requires a pricing formula, expressed as the conditional expectation of a contingent claim under a risk-neutral probability measure Q.

In this paper, we propose a Bayesian MCMC computational approach for calibration of SV models based on combined observations of returns and derivatives. This may be referred to as the hybrid volatility approach. To put our work in perspective, we briefly summarize a number of important papers addressing the aforementioned issue of returns

or/and derivatives.

Aït-Sahalia et al. (1998) considered testing the hypothesis $f_1 = f_2$ using the data of S&P 500 index and the SPX (the Chicago Board Options Exchange) calls and puts, where f_1 represents the (conditional) state-price density (SPD) of the underlying asset price S_T at the option maturity T given the information $\mathcal{F}_t$ collected up to time t, with respect to the risk-neutral probability Q, while f_2 is a version of the same SPD with respect to the real-world probability P via the Girsanov transformation. In conducting the test, f_1 and f_2 were estimated from the option and stock prices respectively. The null hypothesis is equivalent to the complete market assumption. The authors came to a conclusion of rejecting the hypothesis, which implies the market is incomplete. In the context of calibrating parametric volatility models, the above conclusion suggests that the historical and implied volatility approaches appear to complement, rather than duplicate each other. Since a SV model introduces an incomplete market due to the unobservable and nontradable volatility process, it is sensible to use both underlying asset values and option prices as data, and combine historical and implied volatility components.

Avellaneda et al. (1998) formulated model calibration as an optimization problem with constraints. The functional to be minimized is the relative entropy (Kullback-Leibler distance) between a proposed probability model and the empirical distribution of historical data collected on some related economic variables. Each pricing formula was added as a constraint (via Lagrange multipliers). This variational problem was

transformed into an equivalent version of Hamilton-Jacobi-Bellman equations which the authors proposed to solve numerically using finite difference methods. This formulation can be regarded as a Bayesian inference problem, where the likelihood for historical data corresponds to the relative entropy and the prior serves as a regularization factor. An interesting connection to volatility calibration is that every option pricing formula introduces an additional regularization factor, tuning the proposed model toward not only the return data, but also the option data.

Chernov and Ghysels (2000) applied EMM to Heston's SV model [Heston (1993)] for volatility estimation and studied several related problems: option pricing, hedging, and the affinity assumption for the volatility process. The closed-form formula for pricing European call option provided by Heston [also see Bakshi et al. (1997)] paves a road for calculating the implied volatility from the option prices in conjuction with the return data.

A systematic study of SV models via MCMC methods started with an important early work — Jacquier, Polson and Rossi (1994) [denoted by JPR (1994) in what follows, since the version of MCMC we adopt is adapted from that paper and its follow-up JPR (1999)]. Around the basic Log-AR(1) SV model and with only return data used, the authors laid down the ground work for the MCMC in this context. JPR (1999) extended the framework in JPR (1994) to the cases of correlated returns and volatility, heavy-tail data, etc. In spite of the significant contributions made in these two papers, many challenging issues remain unresolved, such as how to choose the proposal density in the

Metropolis dynamics, and how to evaluate the MCMC convergence. A number of papers were written regarding more efficient implementation of MCMC. The most noticeable works were due to Neil Shephard and his collaborators [see Shephard (1996), Shephard and Pitt (1997), Pitt and Shephard (1998), Kim, Shephard and Chib (1998), Chib, Nardari and Shephard (1998), etc.] The application of MCMC to Log-AR(1) model was also a part of Chen (1997). Overall, these works are restricted to using return data, with no derivatives involved.

Jones (2000) appears to be the first one to apply MCMC methods to calibration of SV models using both return and option data. The model considered is an extension of Heston's model, hence no closed-form solution is available for the European call option pricing formula. To make the implied volatility more tractable, Jones proposed the following nice approximation [see Equations (4) and (5) in Jones (2000)]:

$$E_t^Q \left(\frac{1}{\tau} \int_t^{t+\tau} V_s \, ds \right) = A + BV_t,$$

and

$$IV_t = E_t^Q \left(\frac{1}{\tau} \int_t^{t+\tau} V_s \, ds \right) + \epsilon_t,$$

where ϵ_t's are iid $N(0, \lambda^2 IV_{t-1}^2)$ random variables. The first equation assumes that the conditional expectation of the average volatility over the life of the option is an affine function of the current volatility, and the second equation interprets the observed implied volatility as the above expected average volatility perturbed by an error term, whose

variance is proportional to IV_{t-1}^2 — a similar idea to GARCH models. The author was aware of possible misspecification caused by assuming the simplified model, and checked the validity of the assumption.

In this paper we apply MCMC methods to calibration of SV models. Compared to most of the previous works in this area, our MCMC algorithms run simultaneously under the probability measures P and Q, not just under P. Although our approach is close to Jones (2000), we do not use any simple approximations; instead, we resort to direct MCMC and numerical integration to approximate the option pricing formulas for the implied volatility process. This is not intended to downplay the value of ad hoc simplications used by Jones. In practice, we support using simple approximations in all possible situations, accompanied by sound justifications as Jones did, because that often leads to more efficient computation, which is a great contribution made by Jones. The choice of adequate approximations is clearly problem-dependent. Our main purpose in this paper is to demonstrate the applicability of MCMC methodology in complicated stochastic modeling, such as SV models with both return and derivative data included.

MCMC methodology will become more and more useful in empirical finance, in our opinion, due to its versatility and in particular, the "divide-and-conquer" capability — breaking a complex problem into a sequence of simpler subproblems. As the complexity of models and databases in financial markets increases exponentially, it becomes evident that less and less problems enjoy neat analytic solutions. The high price we pay for MCMC is the extreme computational intensity, which limits the popularity of MCMC

in practice for now. Nevertheless, MCMC is promising and will become more applicable with the growing computational power and facilities.

To reduce the computational burden in simulating the dynamics of both real-world and risk-neutral world, we will concentrate on the case of uncorrelated return and volatility processes in this paper, and report the numerical results only using simulated data. The case with correlations between returns and volatility is considerably more challenging and will be studied in another paper, with analysis using return and option data in real markets.

The paper is organized as follows. The mathematical formulation is presented in Section 2. Section 3 demonstrates the MCMC algorithms applied to the special Log-AR(1) model, and reports the numerical results. Section 4 describes certain large-sample properties of posterior distributions of the parameters in SV models, which help us understand the numerical results and also bridge the Bayesian and frequentist approaches.

2 Mathematical framework

We present the general SV model introduced in Romano and Touzi (1997), which will serve as the basis for developing the Bayesian computational method proposed in this work. The model covers those considered in Hull and White (1987), Renault and Touzi (1996), among others. It is a two-factor model induced by two independent sources of noise: $W^S = \{W_t^S,\ t \geq 0\}$ for the underlying asset process $S = \{S_t,\ t \geq 0\}$, and $W^\sigma = \{W_t^\sigma,\ t \geq 0\}$ for the volatility process $\sigma = \{\sigma_t,\ t \geq 0\}$ respectively, and $W = (W^S, W^\sigma)$

is assumed to be a standard 2D Brownian motion defined in a probability space $(\Omega, \mathcal{F}, P)$.
Let $\{\mathcal{F}_t, \ t \geq 0\}$ be the P-augmentation (with null sets added) of the filtration generated
by W. The riskless short interest rate is assumed to be a deterministic function $r_t, t \geq 0$.

The dynamics of the SV model under P is given by the SDEs

$$\frac{dS_t}{S_t} = \mu_t \ dt + \sigma_t \sqrt{1 - \rho_t^2} \ dW_t^S + \sigma_t \rho_t \ dW_t^\sigma \tag{1}$$

$$dh_t = a_t \ dt + b_t \ dW_t^\sigma \tag{2}$$

where $h_t = g(\sigma_t)$ with some one-to-one function g, such as a logarithmic function. In
what follows, both σ and $h = \{h_t, \ t \geq 0\}$ are referred to as volatility processes. See
Renault and Touzi (1996), Romano and Touzi (1997), and Fouque, Papanicolaou and
Sircar (1999) for some needed measurability and regularity conditions on the function
g, the processes $\{\mu_t\}$, $\{a_t\}$, $\{b_t\}$ and $\{\rho_t\}$.

For technical simplicity, a_t, b_t and ρ_t are often assumed to be functions of h_t,
but not of S_t. However, this SV model still incoporates the *leverage effect*, i.e. the
asymmetric effect of asset price change on volatility, due to the inclusion of correla-
tion process $\{\rho_t\}$ between the innovations $\{dW_t^\sigma\}$ to the volatility and the innovations
$\left\{\sqrt{1 - \rho_t^2} \ dW_t^S + \rho_t \ dW_t^\sigma\right\}$ to the returns. See Campbell et al. (1997) p497 for more
details.

The requirement of no arbitrage is essentially equivalent to the existence of an equiva-
lent martingale measure (risk-neutral measure) Q, defined by its Radon-Nikodým deriva-
tive with respect to P [for the Girsanov transformation, see Øksendal (1995) Theorem

9

8.26]:

$$M_t = \left.\frac{dQ}{dP}\right|_{\mathcal{F}_t} = \exp\left(-\int_0^t \lambda_u \ dW_u^S - \frac{1}{2}\int_0^t \lambda_u^2 \ du - \int_0^t \nu_u \ dW_u^\sigma - \frac{1}{2}\int_0^t \nu_u^2 \ du\right), \quad (3)$$

where the two processes $\lambda = \{\lambda_t, \ t \in [0,T]\}$ and $\nu = \{\nu_t, \ t \in [0,T]\}$ are interpreted as the asset risk premium and the volatility risk premium relative to the two uncertainty sources W^S and W^σ. It will be verified in Appendix that a risk-neutral measure Q exists if and only if

$$\sigma_t \ \left(\lambda_t\sqrt{1-\rho_t^2} + \nu_t\rho_t\right) = \mu_t - r_t, \tag{4}$$

for a.e. $t \in [0,T]$ and P-a.s.

The SV model corresponds to an incomplete market in which the volatility σ is neither observable nor tradable. This also implies the non-uniqueness of martingale measure Q. Looking from another angle, the relationship (4) does not fix the volatility risk premium ν. Instead, each choice of ν determines a corresponding measure $Q(\nu)$ via (3). In other words, there is a family of martingale measures parameterized by the premium ν. Each member $Q(\nu)$ is called an admissible martingale measure in Romano and Touzi (1997). Unless necessary, we simply write Q for $Q(\nu)$.

It can be verified via Itô's formula that

$$S_t = S_0 \ \exp\left[\int_0^t (\mu_u - \sigma_u^2/2) \ du + \int_0^t \sigma_u\sqrt{1-\rho_u^2} \ dW_u^S + \int_0^t \sigma_u\rho_u \ dW_u^\sigma\right] \tag{5}$$

satisfies (1). Let

$$\widetilde{W}_t^S = W_t^S + \int_0^t \lambda_u \, du \tag{6}$$

$$\widetilde{W}_t^\sigma = W_t^\sigma + \int_0^t \nu_u \, du. \tag{7}$$

The dynamics of the SV model under Q is given by

$$\frac{dS_t}{S_t} = r_t \, dt + \sigma_t \sqrt{1 - \rho_t^2} \, d\widetilde{W}_t^S + \sigma_t \rho_t \, d\widetilde{W}_t^\sigma \tag{8}$$

$$dh_t = (a_t - b_t \nu_t) \, dt + b_t \, d\widetilde{W}_t^\sigma. \tag{9}$$

Therefore,

$$S_T = S_t \, \exp \left[\int_t^T (r_u - \sigma_u^2/2) \, du + \int_t^T \sigma_u \sqrt{1 - \rho_u^2} \, d\widetilde{W}_u^S + \int_t^T \sigma_u \rho_u \, d\widetilde{W}_u^\sigma \right]$$

$$= S_t \, e^{\eta_{t,T}} \, \exp \left(\int_t^T \sigma_u \sqrt{1 - \rho_u^2} \, d\widetilde{W}_u^S \right), \tag{10}$$

where

$$\eta_{t,T} = \int_t^T (r_u - \sigma_u^2/2) \, du + \int_t^T \sigma_u \rho_u \, d\widetilde{W}_u^\sigma.$$

In this paper, we assume the volatility risk premium ν to be a constant parameter and fit it by observed option prices. In general, assume that for $t \in [0, T]$, ν_t depends only on h_t, but not S_t. It follows from (9) that the process h (or σ) is independent of $\widetilde{W}^S$. Hence conditioning on $\mathcal{F}_t$ and $\{\sigma_u, \ u \in [t, T]\}$,

$$\log \frac{S_T}{S_t e^{\eta_{t,T}}} \sim \text{Normal} \left(0, \int_t^T \sigma_u^2 (1 - \rho_u^2) \, du \right). \tag{11}$$

11

We denote the conditional expectation operators under Q by

$$E_t(\ \cdot\ ) = E_Q(\ \cdot\ |\ \mathcal{F}_t\ )$$

and

$$E_{t,\sigma}(\ \cdot\ ) = E_Q(\ \cdot\ |\ \mathcal{F}_t,\ \sigma_u, u \in [t, T]\ ).$$

For a (European) contingent claim $f(S_T)$ based on the asset price S_T at maturity T, its time t price is given by

$$
\begin{aligned}
V_t &= E_t \left[e^{-\int_t^T r_u du} f(S_T) \right] \\
&= E_t \left[e^{-\int_t^T r_u du} E_{t,\sigma} f(S_T) \right] \\
&= E_t \left[e^{-\int_t^T r_u du} \int_{-\infty}^{\infty} f \left(S_t e^{x+\eta_{t,T}} \right)\ p(x)\ dx \right],
\end{aligned}
\tag{12}
$$

where $p(x)$ is the normal density with mean zero and variance $\int_t^T \sigma_u^2 (1 - \rho_u^2) du$. In particular, for the European call option with $f(S_T) = (S_T - K)^+$, we have

$$
V_t = E_t \left[e^{-\int_t^T r_u du} \int_{\log \frac{K}{S_t} - \eta_{t,T}}^{\infty} \left(S_t e^{x+\eta_{t,T}} - K \right)\ p(x)\ dx \right].
\tag{13}
$$

Remarks:

- The dynamics of historical volatility follows either (2) (P-dynamics, the real world) or (9) (Q-dynamics, the risk-neutral world). Hence the Bayesian calibration of historical volatility is carried out by simulating from the posterior distribution of $\{\sigma_u,\ u \in [0, t]\}$ (or equivalently $\{h_u,\ u \in [0, t]\}$ given the observation $\{S_u,\ u \in$

12

$[0, t]\}$. The posterior distribution under either P or Q can be used, although the P-dynamics (2) is more preferable.

- The notion of implied volatility for this SV model goes beyond the original Black-Scholes sense, because the pricing formula (12) cannot be inverted except for rare cases, e.g. the Black-Scholes or Heston's model. Besides the historical SV model (1)–(2) and the observed asset data, there is a new parameter ν in (9) which determines which martingale measure Q governs the risk-neutral dynamics, and can be fitted by the observed derivative data. This means "option prices complete the market in the SV model". There are a number of ways to interwave the volatility risk premium ν and the derivative price V_t, and to add an implied component to the historical volatility h. Some of them amount to adding a noise term to h_t [as in Jones (2000)] or V_t [as in Jacquier and Jarrow (2000)]. In this paper, we will treat the observed derivative price as a perturbed version of V_t. A necessary and demanding step in our algorithms is to perform numerical integration in (12) which involves Monte Carlo simulation of the latent volatility path $\{h_u, \ u \in [t, T]\}$ under the Q-dynamics (9). Note a distinction between the pursuit of implied volatility and its historical counterpart: given the information $\mathcal{F}_t$, to simulate the "future" volatility $\{h_u, \ u \in [t, T]\}$ via (9) is a prediction or extrapolation problem (as needed for implied volatility); while the simulation of $\{h_u, \ u \in [0, t]\}$ via (2) is an estimation or interpolation problem (as needed for historical volatility). Details will be given in Section 3.

13

Discretization

Discretization of (1)-(2) and (8)-(9) is needed before we implement the MCMC procedures. The following standard Euler approximation is adopted here.

Let $\Delta t > 0$ be a fixed mesh — a small time increment. To avoid introducing too many different symbols, we abuse the notation and give some previous symbols new meanings. For example, we simply denote $S_{n\Delta t}$ by S_n, and make similar changes in subscript for other processes. The discrete-time version of (1)-(2) becomes

$$\frac{S_{n+1} - S_n}{S_n} = \mu_n \, \Delta t + \sigma_n \sqrt{1 - \rho_n^2} \, \sqrt{\Delta t} \, \epsilon_{n+1}^{(1)} + \sigma_n \rho_n \, \sqrt{\Delta t} \, \epsilon_{n+1}^{(2)} \tag{14}$$

$$h_{n+1} - h_n = a_n \, \Delta t + b_n \, \sqrt{\Delta t} \, \epsilon_{n+1}^{(2)}, \tag{15}$$

where

$$\epsilon_{n+1}^{(1)} = (\Delta t)^{-1/2} \left[W_{(n+1)\Delta t}^S - W_{n\Delta t}^S \right] \text{ and } \epsilon_{n+1}^{(2)} = (\Delta t)^{-1/2} \left[W_{(n+1)\Delta t}^\sigma - W_{n\Delta t}^\sigma \right],$$

thus $\epsilon_n^{(1)}$, $\epsilon_n^{(2)}$, $n = 0, 1, 2, \ldots$ are iid $N(0,1)$ random variables under P. Simailarly, the discrete-time version of (8)-(9) becomes

$$\frac{S_{n+1} - S_n}{S_n} = r_n \, \Delta t + \sigma_n \sqrt{1 - \rho_n^2} \, \sqrt{\Delta t} \, \epsilon_{n+1}^{(3)} + \sigma_n \rho_n \, \sqrt{\Delta t} \, \epsilon_{n+1}^{(4)} \tag{16}$$

$$h_{n+1} - h_n = (a_n - b_n \nu_n) \, \Delta t + b_n \, \sqrt{\Delta t} \, \epsilon_{n+1}^{(4)}, \tag{17}$$

where

$$\epsilon_{n+1}^{(3)} = (\Delta t)^{-1/2} \left[\widetilde{W}_{(n+1)\Delta t}^S - \widetilde{W}_{n\Delta t}^S \right] \text{ and } \epsilon_{n+1}^{(4)} = (\Delta t)^{-1/2} \left[\widetilde{W}_{(n+1)\Delta t}^\sigma - \widetilde{W}_{n\Delta t}^\sigma \right],$$

thus $\epsilon_n^{(3)}$, $\epsilon_n^{(4)}$, $n = 0, 1, 2, \ldots$ are iid $N(0,1)$ random variables under Q.

3 MCMC algorithms and numerical results

In this section, we present MCMC algorithms for calibration of a hybrid SV model, which combines the historical volatility based on the Log-AR(1) model and the implied volatility estimated from European call option prices.

The Log-AR(1) model, expressed as

$$Y_n = e^{h_n/2} \left(\sqrt{1 - \rho^2}\ \epsilon_n^{(1)} + \rho\ \epsilon_n^{(2)} \right) \tag{18}$$

$$h_n = \alpha + \beta\ h_{n-1} + \eta\ \epsilon_n^{(2)}, \tag{19}$$

is a special case of (14)-(15) by estimating and removing the drift $\mu_n \Delta t$ in (14) and letting

$$\frac{S_n - S_{n-1}}{S_{n-1}} - \mu_{n-1}\Delta t = Y_n, \quad \sigma_n^2 \Delta t = e^{h_n},$$

$$\rho_n = \rho, \quad a_n \Delta t = \alpha + (\beta - 1)h_n, \quad \text{and} \quad b_n \sqrt{\Delta t} = \eta$$

in (14)-(15).

Later when adding implied volatility to the model, we will assume the volatility risk premium $\nu_n = \nu$, and express (16)-(17) accordingly as

$$Y_n = -(\mu_{n-1} - r_{n-1})\ \Delta t + e^{h_n/2} \left(\sqrt{1 - \rho^2}\ \epsilon_n^{(3)} + \rho\ \epsilon_n^{(4)} \right) \tag{20}$$

$$h_n = (\alpha - \nu\ \eta\ \Delta t) + \beta\ h_{n-1} + \eta\ \epsilon_n^{(4)}. \tag{21}$$

Furthermore, our study in this paper is restricted to the uncorrelated case with

$\rho = 0$. We will discuss the MCMC volatility calibration in two separate subsections for the historical volatility and hybrid volatility respectively.

3.1 Historical volatility approach

3.1.1 Algorithm

Our study of MCMC consists of a "top-down" path and a "bottom-up" path, both linked to the basic Log-AR(1) model

$$Y_t = e^{h_t/2} \epsilon_t^{(1)} \tag{22}$$

$$h_t = \alpha + \beta h_{t-1} + \eta \epsilon_t^{(2)}. \tag{23}$$

Note that instead of n, here we denote the discrete time by $t = 1, \ldots, T$.

The top-down path generates a synthetic return data set $y = (y_1, \ldots, y_T)$ by (i) assigning initial values to parameter $\theta = (\alpha, \beta, \eta)$, (ii) simulating the volatility sequence $h = (h_1, \ldots, h_T)$ via (23), (iii) simulating y via (22) conditioning on h. Steps (ii) and (iii) are direct simulation with no MCMC involved.

The bottom-up path aims at several posterior distributions: $p(\theta|y)$ as a basis for the Bayesian inference on θ, and $p(h|y)$ for generating realizations of h which make option pricing possible. Both $p(\theta|y)$ and $p(h|y)$ are marginals of the (joint) posterior distribution $p(\theta, h|y)$. However, $p(\theta, h|y)$ is not only analytically intractable, but also difficult to perform direct Monte Carlo simulation. The basic idea of MCMC is to cycle through the conditional distributions $p(\theta|h, y)$ and $p(h|\theta, y)$ iteratively. It is easy to

16

simulate from $p(\theta|h, y)$, but much harder to generate samples from $p(h|\theta, y)$ due to high dimensionality of h. JPR (1994) tackled the problem by decomposing $p(h|\theta, y)$ further into a set of conditional distributions $p(h_t|h_{-t}, \theta, y)$ and cycle through them, where h_{-t} denotes the rest of the h vector other than h_t. Following the Markov property, the long vector h_{-t} in the conditioning can be replaced by (h_{t-1}, h_{t+1}) (the immediate past and future), thus

$$p(h_t|h_{-t}, \theta, y) \propto p(y_t|h_t) \; p(h_t|h_{t-1}, \theta) \; p(h_{t+1}|h_t, \theta).$$

Metropolis algorithms (a special version of MCMC) allow us to generate MCMC samples without dealing with those complicated normalization constants. For the SV models we considered, including the Log-AR(1), the convergence of Metropolis algorithms can be verified by invoking the general theory given in Tierney (1994).

The MCMC algorithm proposed in JPR (1994) (1999) can be summarized in the following steps:

Step 1   Initialize h and θ.

Step 2   Sample h_t from $p(h_t|h_{-t}, \theta, y)$ for $t = 1, \ldots, T$.

Step 3   Sample θ from $p(\theta|h, y)$.

Step 4   Go back to Step 2 and continue.

The algorithm above is usually quoted as a single-move sampler because the components of h process are updated one by one sequentially. Each iteration of the Metropolis algorithm consists of updating T components of h, and it will run a large number N

iterations before the sampler terminates.

Step 3 follows the standard Bayesian linear model theory [Gelman et al. (1998), Chapter 8]: starting from a non-informative conjugate prior for θ

$$p(\alpha, \beta, \eta) \propto \eta^{-2}$$

and observing that θ and y are conditionally independent given h, the posterior for θ given h is given by

$$p(\theta|h) = \mathcal{N}_2(\mu, \ (H^\top H)^{-1}\eta^2) \ IG((T-2)/2, \ Ts^2/2), \tag{24}$$

where $\mathcal{N}_k$ denotes the k-variate normal distribution, and IG the inverse gamma distribution; $\mu = (H^\top H)^{-1} H^\top h_{2,T}$ with $h_{2,T} = (h_2, \ldots, h_T)^\top$ (the notation $A^\top$ represents the transpose of matrix A), H is an $(T-1) \times 2$ matrix whose row t is $(1, h_t)$, $t = 1, \ldots, T-1$, and Ts^2 is the standard sum of squared errors.

In Step 2, we need to simulate, for each $t = 1, \ldots, T$, from the single-site conditional distribution $p(h_t|h_{t-1}, h_{t+1}, \theta, y)$, expressed as

$$p(h_t|h_{t-1}, h_{t+1}, \theta, y) \propto \exp\left\{-\frac{1}{2}\left[h_t + y_t^2/e^{h_t} + (h_t - m_t)^2/b^2\right]\right\}, \tag{25}$$

where $m_t = [\alpha(1-\beta) + \beta \ (h_{t-1} + h_{t+1})]/(1+\beta^2)$ and $b^2 = \eta^2/(1+\beta^2)$.

Applying the Metropolis algorithm locally at each t to (25), JPR (1994) (1999) proposed an inverse gamma blanketing density (proposal density) q to approximate the (log-normal) density of e^{h_t} in (25). Our simulation study, however, found that that

18

proposal often results in misspecified values of h_t when the true value of η is near zero. More importantly, it is inconvenient to apply that approximation scheme to the hybrid volatility later. Denote the updated value of h after iteration i by $h^{(i)}$ and assume that the current value within iteration $i+1$ is $(h_1^{(i+1)}, \ldots, h_{t-1}^{(i+1)}, h_t^{(i)}, \ldots, h_T^{(i)})$. We choose the blanketing density to be $q(h_t^{(i+1)}|h_t^{(i)}) \sim N(h_t^{(i)}, c_0\hat{\eta})$, where $\hat{\eta}$ is an estimate of η obtained so far and c_0 is a constant chosen judiciously, e.g. $c_0 = 0.1$ in our study. Note that this blanketing kernel is symmetric, i.e. $q(z|z') = q(z'|z)$ for all $z, z' \in \mathbb{R}$. Therefore, $h_t^{(i+1)}$ sampled from the above q is accepted with probability $\min\left(\frac{p(h_t^{(i+1)}|h_{t-1}^{(i+1)}, h_{t+1}^{(i)}, \theta^{(i)}, y)}{p(h_t^{(i)}|h_{t-1}^{(i+1)}, h_{t+1}^{(i)}, \theta^{(i)}, y)}, 1\right)$.

This modification of Metropolis algorithm/Gibbs sampler works quite well in the current situation.

3.1.2 Results and plots

We simulated a data set with sample size T=500 and initial values $\alpha = -0.8$, $\beta = 0.9$ and $\eta = 0.6$ respectively. After 110,000 MCMC iterations, we discarded the first 10,000 iterations to reduce the initial value effect. Figure 1 shows the simulation trajectory plots, histogram plots and the autocorrelation plots of the sampled β and η. The reason we skipped the plots for α is that they look very similar to those for β, and that the estimation of β has a greater impact thus requires greater accuracy.

Some numerical results are summarized in Table 1. The numbers in parentheses of the first column are true parameter values, the numbers in the next two columns are the posterior means and standard deviations calculated based on the 100,000 iterations.

The real computational time for the MCMC using Matlab 5.2 is approximately four hours on a Pentium III 500MHZ machine.

3.1.3   Diagnostics of MCMC convergence

The general theory of MCMC convergence given in Tierney (1994) covers various MCMC algorithms, including Metropolis dynamics and Gibbs sampler. However, guidelines for diagnostics of MCMC convergence are needed in practice. One thing worthy of notice is the high autocorrelations in the MCMC samples. The acf plots in Figure 1 show significant autocorrelations for β at 5,000 iterations, and for η at almost 20,000 iterations. This is typical when using single-move MCMC methods which update a single component of volatility series at a time, and result in very slow mixing in the samples. Some multi-move MCMC methods were proposed, e.g. in Carter and Kohn (1994), Kim, Shephard and Chib (1998), to correct the slow convergence by sampling a block of highly correlated components simultaneously. But those methods usually rely on some ad hoc approximations which need not be available when we combine hostorical and implied volatilities. Therefore, we stick to single-move MCMC methods in this paper and perform a convergence diagnostics using the criterion given in Gelman and Rubin (1992).

Gelman and Rubin (1992) proposed to run multiple MCMC chains with different initial values sampled from a distribution and monitor the MCMC convergence by calculating a ratio of two different variances, the average variance within a single chain

("within-chain variance") and the average variance cross multiple chains ("between-chain variance"). For any quantity of interest Z, let $Z_m^{(n)}$ be its value in the nth iteration of the mth chain, $n = 1, \ldots, N$ and $m = 1, \ldots, M$. Moreover, define

$$\bar{Z}_m = \frac{1}{N} \sum_{n=1}^{N} Z_m^{(n)}, \quad \bar{Z} = \frac{1}{M} \sum_{m=1}^{M} \bar{Z}_m,$$

$$W_N = \frac{1}{M} \sum_{m=1}^{M} s_m^2 = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{N} \sum_{n=1}^{N} (Z_m^{(n)} - \bar{Z}_m)^2,$$

and

$$B_N = \frac{1}{M} \sum_{m=1}^{M} (\bar{Z}_m - \bar{Z})^2.$$

The quantities W_N and B_N represent the average within-chain and between-chain variances respectively. An estimator of the posterior variance of Z is a sum of W_N and B_N, i.e.

$$\hat{\sigma}_N^2 = \frac{N-1}{N} W_N + B_N.$$

Furthermore, let

$$\hat{V}_N = \hat{\sigma}_N^2 + \frac{B_N}{M}, \quad \nu_N = \frac{2\hat{V}_N^2}{W_N},$$

$$w_N = \frac{1}{M^2} \left[\sum_{m=1}^{M} s_m^4 - \frac{1}{M} \left(\sum_{m=1}^{M} s_m^2 \right)^2 \right],$$

and

$$R_N = \frac{\hat{V}_N}{W_N} \frac{\nu_N}{\nu_N - 2} = \left(\frac{N-1}{N} + \frac{M+1}{M} \frac{B_N}{W_N} \right) \frac{\nu_N}{\nu_N - 2}. \tag{26}$$

21

Gelman and Rubin (1992) recommended using the shrink factor R_N as a criterion to monitor MCMC convergence and stopping the MCMC sampling when R_N is close to 1 with large N. To conduct the test, the approximate distribution of R_N can be derived from an approximate F-distribution, i.e.

$$\frac{NB_N}{W_N} \sim F\left(M - 1, \ 2W_N^2/w_N\right).$$

We take $Z = \beta$ as an example in our study and run 8 independent chains with different initial values selected from a sample distribution, each chain running up to 110,000 iterations with the first 10,000 iterations discarded. Figure 2 contains a plot of the shrink factor R_N against N. As is observed from the plot, the MCMC algorithm converges after a few thousand iterations.

3.1.4 MCMC output of volatility h

Besides an estimate of the parameter θ, Bayesian MCMC methods also deliver volatility paths h as a byproduct. This makes the MCMC approach more useful since h is an important factor in pricing derivatives.

In each of the 110,000 MCMC iterations, a sample of $h = (h_1, \ldots, h_T)$ is obtained. We calculate the posterior mean and standard deviation of h_t at each t and plot them in Figure 3. In the top plot, the blue curve represents the generated (log-volatility) h series, and the red curve is the posterior mean series calculated from the MCMC simulation. In the bottom plot, the dotted curves display a 95% "confidence band" for h. Note that it is a bundle of pointwise (at each t) confidence intervals, not a genuine confidence band

in the sense of simultaneous inference.

These summary plots demonstrate that Bayesian MCMC methods also serve as a smoother in volatility estimation.

3.1.5   Impact of β on MCMC

A noticeable drawback of frequentist methods for calibration of SV models is that they perform poorly when the slope β is close to 1 (clustering effect), which happens to be common for many financial time series. In contrast, MCMC methods work better in this situation. To illustrate this, we generate three different data sets with the same sample size 500 and true parameter $\eta = 0.6$, but different true β values 0.7, 0.8 and 0.9 respectively. The values of α are set accordingly to keep the h series in approximately the same scale. The Bayesian inference procedure is applied to each of three data sets. The histogram estimate of the posterior distribution $p(\beta|y)$ produced by the MCMC procedure looks more desirable for the case with true β value 0.9 than for the other two cases. See Figure 4 for the plots and Table 2 for the numerical summary. Note that here "more desirable" means (i) the shape of $p(\beta|y)$ density is closer to a normal curve; (ii) the posterior mean is closer to the true β value; and (iii) the posterior standard deviation is smaller. A theoretical back-up of these observations is the asymptotic normality proved in Section 4.

3.1.6 Impact of η on MCMC

JPR(1994) provided some empirical evidence for the performance of QML and concluded that as the true η value decreases, the QML procedure has poor sampling properties. Our study leads to the same conclusion for MCMC methods. A possible explanation for this phenomenon is that for very small η values, the adjacent volatility components become highly correlated, which makes the already slow mixing of MCMC samplers even slower.

We run the top-down simulation with the assigned value $\eta = 0.2$. The estimation through the bottom-up path looks less accurate than the case with $\eta = 0.6$. The result is given in Table 3.

3.1.7 Impact of sample size T on MCMC

Another important factor that affects the performance of MCMC procedures is the sample size T. This is demonstrated by running MCMC on four different synthetic data sets with sample sizes 100, 200, 300 and 500. Figure 5 show the four histogram estimates of $p(\beta|y)$, and Table 4 summarizes the numerical results. These results clearly demonstrate that the performance of MCMC depends on the sample size. While the posterior distribution for $T = 500$ nicely peaks around the true value $\beta = 0.9$, the results for $T = 100$ and $T = 200$ are only marginally acceptable where the posterior distributions are left-skewed and have greater standard errors. Again, see Section 4 for the related asymptotic normality of posterior distributions.

The impact of sample size T will be sicussed again in Section 3.2.3 when derivative data are added.

3.2 Hybrid volatility approach combining historical volatility and implied volatility

The historical volatility approach mirrors a standard practice in statistics and econometrics. It is more robust than the implied volatility approach, because it is based on the *past* information on the underlying assets, not tied with any particular derivatives. On the contrary, the implied volatility approach is ad hoc and less stable, because it requires the knowledge of some specific pricing formulas. Denote the time t price of an European call option by C_t. The connection between the implied volatility and the option price C_t has an implicit predictive feature, since C_t is expressed as a risk-neutral average of *future* volatility and prices in the market during the residual lifetime of the option in $(t, t + \tau]$. Although we do not advocate using implied volatility to replace historical volatility, we do suggest adding the implied volatility component when option pricing is relevant.

In this subsection, we propose a modified MCMC algorithm for hybrid volatility in which both return data and option prices are used to help improve the calibration.

3.2.1 Algorithm

Assume the observations include the return data y as well as the option prices $C = (C_t, \; t = 1, \ldots, T)$. Recall that in the equation (13), V_t denotes the call option price at

time t, expressed as the conditional expectation with respect to a risk-neutral measure. To make a subtle distinction, we assume that the observed price C_t is a perturbed version of V_t (to be explained later).

The hybrid volatility model that drives the new MCMC algorithm consists of

Part 1: the same historical volatility model Log-AR(1) represented by (22) and (23); and

Part 2: the pricing formula [a special version of (13)]

$$V_t = e^{-r\tau} \; E_t \left[\int_{\log \frac{K}{S_t} - \eta_{t,t+\tau}}^{\infty} \left(S_t e^{x + \eta_{t,t+\tau}} - K \right) \; p(x) \; dx \right] \tag{27}$$

for the implied volatility, where

$$\eta_{t,t+\tau} = \int_t^{t+\tau} \left(r - e^{h_u}/2 \right) \; du,$$

with the strike price K, the maturity $t + \tau$ for the option, and the density $p(x)$ of $N\left(0, \; \int_t^{t+\tau} e^{h_u} du \right)$; also, assume the additional observation equation

$$\log(C_t/V_t) \;\; = \;\; \delta \; \epsilon_t^{(3)}, \tag{28}$$

where $\delta > 0$ is a new parameter, and $\{\epsilon_t^{(3)}, \; t = 1, 2, \ldots\}$ is a sequence of iid $N(0,1)$ random variables, independent of $\{\epsilon_t^{(1)}\}$ and $\{\epsilon_t^{(2)}\}$.

Part 1 needs no further explanation. In Part 2, the computation needed for V_t makes the new MCMC algorithm much slower than the one presented in Section 3.1. The numerical integration in V_t consists of two steps. Step 1 is the integration inside the

conditional expectation $E_t(\cdot)$. Step 2 is the conditional expectation itself. For Step 1, given a realization of $(h_t, \ldots, h_{t+\tau})$, we generate iid samples of X from $p(x)$ and use an ergodic average to approximate the integral. Step 2 is to generate the future path $(h_{t+1}, \ldots, h_{t+\tau})$ given the information up to time t, which should not be driven by the real-world dynamics (23) but by the risk-neutral dynamics [a special version of (17)]

$$h_{t+1} = (\alpha - \nu\eta) + \beta \, h_t + \eta \, \epsilon_{t+1}^{(4)}. \tag{29}$$

Having generated N future paths of $(h_{t+1}, \ldots, h_{t+\tau})$, V_t can be approximated by the sample average of N corresponding integrals calculated in Step 1.

The proposal (28) might be controversial because it seems to suggest the inability of pricing an option in general. We would argue that V_t should be the correct price without errors *if* all quantities in (27) can be fully specified. However, that is unrealistic due to the incomplete market and the nonuniqueness of risk-neutral measures. Each risk-neutral measure $Q(\nu)$ is parameterized by a given value of the unknown ν. The observed market price C_t contains the information regarding a specific value of ν, and we need C_t to identify the parameter ν. Therefore, the corruption from the expected price V_t to the observed price C_t is inevitable and appears to be a reasonable assumption. Define $L_t = \log(C_t/V_t)$. Intuitively, the assumption $L_t \sim N(0, \delta^2)$ means that percentage wise, the chance of over-pricing an option is the same as under-pricing it.

The augmented parameter is $\theta = (\alpha, \beta, \eta, \nu, \delta)$. The new MCMC procedure will operate with the two conditional distributions $p(\theta|h, y, C)$ and $p(h|\theta, y, C)$.

Sampling from $p(\theta|h, y, C)$

A closer look at the model (22), (23), (27) and (28) suggests the following factorization based on conditional independence

$$p(\theta|h, y, C) = p(\alpha, \beta, \eta|h) \; p(\nu, \delta|h, y, C),$$

in which the first factor can be handled by the MCMC method used in Section 3.1. To simulate from the second factor, we will alternate between $p(\delta|\nu, h, y, C)$ and $p(\nu|\delta, h, y, C)$. Write $V = (V_1, \ldots, V_T)$, we have

$$p(\delta|\nu, h, y, C) = \; p(\delta|V, C) \propto \; p(\delta) \; \frac{1}{\delta^T} \; \exp\left(-\sum_{t=1}^{T} \frac{L_t^2}{2\delta^2}\right), \tag{30}$$

and

$$p(\nu|\delta, h, y, C) \; \propto \; p(\nu) \; p(h|\nu) \; p(V|\nu, h) \; p(C|\delta, V) \tag{31}$$

$$\propto \; p(\nu) \; p(h|\nu),$$

because $p(V|\nu, h) = 1$ and $p(C|\delta, V)$ does not involve ν.

In (30), L_t is not a function of δ. A conjugate inverse gamma prior $\delta \sim IG(v_0, v_0 s_0^2)$ leads to the posterior

$$p(\delta|\nu, h, y, C) \sim IG\left(v_0 + T - 1, \; v_0 s_0^2 + \sum_{t=1}^{T} L_t^2\right).$$

To make the prior less informative, we set $v_0 = 5$ to be much smaller than T and $v_0 s_0^2 = 0.002$ so that the order of magnitude is smaller than $\sum_{t=1}^{T} L_t^2$.

28

In (31), a normal conjugate prior for ν along with the Gaussian linear model (29) will imply a normal posterior equivalent to $p(\nu|\delta, h, y, C)$.

Sampling from $p(h|\theta, y, C)$

Modifying (25), we sample from $p(h|\theta, y, C)$ by cycling through the single-site conditional distributions

$$p(h_t|h_{t-1}, h_{t+1}, \theta, y, C) \propto \exp\left\{-\frac{1}{2}\left[h_t + y_t^2/e^{h_t} + (h_t - m_t)^2/b^2 + \frac{L_t^2}{\delta^2}\right]\right\}. \quad (32)$$

Compared to the MCMC algorithm given in Section 3.1, the new MCMC contains more steps:

Step 1 Initialize h and θ.

Step 2.1 For each t, generate the path $(h_{t+1}, \ldots, h_{t+\tau})$ via (29), then calculate V_t and L_t.

Step 2.2 Update the value of h_t by sampling from (32) with the corresponding Metropolis algorithm.

Step 3 Sample θ from $p(\theta|h, y, C)$. Apply the direct Bayesian inference with conjugate priors to $p(\alpha, \beta, \eta|h)$ and $p(\nu, \delta|h, C)$ separately.

Step 4 Go back to Step 2.1 and continue.

The major changes are the new Step 2.1, and the addition of option data C to the conditional distributions in Step 2.2 and Step 3.

3.2.2 Results and plots

For $T = 500$, we generate the volatility series h and return data y as before using (22) and (23). At each t, we also produce an at-the-money call option with strike price $K = S_t$ and maturity $t + 30$ which makes the option mature in a month. V_t is calculated using the approximation scheme described in Section 3.2.1, and C_t is further generated using formula (28). Having completed this top-down path of synthesis, we go through the bottom-path by running the modified MCMC (also presented in Section 3.2.1).

The computational time for this new MCMC is considerably longer than the previous MCMC, primarily due to the numerical integration for V_t. With each independent initialization seed, we ran a total of 55,000 MCMC iterations and discarded the first 5,000 iterations. The Matlab program (Version 5.2) took approximately 40 hours on a Pentium III 500MHZ machine.

Table 5 presents the numerical results.

3.2.3 Comparisons with the historical volatility approach

The intensive MCMC computation for the hybrid volatility approach is not wasteful. The accuracy of estimating both the parameter θ and the volatility series h is improved dramatically.

Improvement in parameter estimation

Comparing Table 5 to Table 1, we can see that the hybrid volatility approach re-

duces the posterior standard deviation of β. In order to see more clearly how the hybrid volatility approach can improve the estimation, we will use examples with smaller sample sizes for illustration. Four samples of different sizes, $T = 50, 100, 200, 500$, are used to run the MCMC simulations with the hybrid volatility. Table 6 summarizes the results of numerical comparison between the hybrid volatility approach and the historical approach. The corresponding histogram plots of β are displayed in Figure 6.

Overall, adding option data amounts to increasing the effective sample size. The improvement seems more significant in the cases with small sample sizes. Notice how well the nearly bell-shaped histogram looks even for the case $T = 50$ when using the hybrid volatility.

Improvement in volatility estimation

Our study demonstrates that the MCMC hybrid volatility approach also delivers more accurate estimates for the volatility series.

To illustrate this, we compare the results from the hybrid model and the historical model for $T = 500$. In the two plots in Figure 7, the blue curves represent the original log-volatility time series h we generated, and the red curves represent the posterior means of h computed from the MCMC simulation. The top and bottom plots correspond to the historical and hybrid volatilities respectively.

The mean square errors (MSE) for the estimation are calculated as a numerical measure of estimation accuracy. Our calculation yields the MSE 0.6206 for the historical

volatility approach, and the MSE 0.3392 for the hybrid volatility approach.

3.2.4 Volatility smile in SV models

Hull and White (1987) is an early study for option pricing in SV models and related volatility smile. Renault and Touzi (1996) provided a theoretical justification. Another byproduct of our proposed MCMC methods is to show some numerical evidence of volatilty smile in the SV models.

Repeat the same top-down path from θ to h to y and C, but this time we generate call option prices with the same maturity but different strike prices K. The resulting Black-Scholes implied volatility against K can be plotted. In doing this, we set $\alpha = -0.8, \beta = 0.9, \eta = 0.6$ and $\nu = 0.1$. At time $t = 100$, we observe the stock price $S_t = \$119.63$ and the volatility $\$0.0236$. Define ten call options with K evenly spread out from $\$100$ to $\$145$ and the same maturity 60 days, we obtain the B-S implied volatility for each option. The option prices and corresponding B-S implied volatilities are reported in Table 7.

Figure 8 is a plot of the calculated B-S implied volatilities against the strike prices. The plot shows greater volatilities associated with high (out-of-the-money) or low (in-the-money) strike prices, and smaller volatilities associated with strike prices close to the stock price (at-the-money). If the real market is driven by a SV model, then the observed option prices will also be driven in a complicated manner. Nevertheless, if we ignore the underlying model and always use Black-Scholes model to calculate the implied

volatility, we will observe the volatility smile as illustrated here.

4   Large-sample properties of posterior distributions

To help the reader, we begin with a list of (old and new) notation used in this section.

- $Y^n = (Y_1, \ldots, Y_n)$: a finite sample (of size n) of the discrete-time return process Y, and $y^n = (y_1, \ldots, y_n)$ for the realization y similarly; (Note: unless necessary, we do not always make such a notational distinction between a random object and its realizations.)

- $h^n = (h_1, \ldots, h_n)$: a finite sample of h;

- $x^n = (y^n; \ h^n)$: augmented data;

- $\theta \in \Theta$: a finite-dimensional parameter (column) vector, e.g. $\theta = (\alpha, \beta, \eta, \rho, \nu)^\top$ in the Log-AR(1) model;

- $f_{x^n}(\cdot|\theta)$: the conditional density of x^n given θ, thus $f_{x^n}(x^n|\theta)$ and $f_\theta(\theta|x^n)$ denote the likelihood and the posterior density respectively, with the augmented data;

- $l_{x^n}(\theta) = \log f_{x^n}(x^n|\theta)$: log-likelihood given x^n;

- $\hat{\theta}(x^n)$: maximum likelihood estimate (MLE) of θ based on x^n;

- $\nabla^2 l_{x^n}(\theta_0)$: Hessian matrix of the likelihood given x^n evaluated at $\theta = \theta_0$;

- $\Sigma_{x^n} = [-\nabla^2 l_{x^n}(\hat{\theta}(x^n))]^{-1}$: the inverse of the empirical Fisher information matrix evaluated at the MLE $\hat{\theta}(x^n)$ (if exists);

- Given an observation y^n, the actual posterior density $f_\theta(\theta|y^n)$, log-likelihood $l_{y^n}(\theta)$, MLE $\hat{\theta}(y^n)$ and empirical Fisher information matrix $-\nabla^2 l_{y^n}(\hat{\theta}(y^n))$, etc. can be defined similarly.

Our study in this section concerns asymptotic properties of the proposed Bayesian methods, such as consistency and asymptotic normality of the posterior distribution $p(\theta|y^n)$ in SV models, as the sample size n goes to infinity. As a basis for Bayesian inference, the expression of posterior $p(\theta|y^n)$ often involves complex integrals over the latent variable h, which hinders direct computation. The key step is to write $p(\theta|y^n)$ as a mixture of intermediate conditional distributions $p(\theta|x^n)$ given the augmented data. This enables us to covert the large-sample properties of $p(\theta|x^n)$ to those for $p(\theta|y^n)$.

The consistency of the MLE $\hat{\theta}(x^n)$ is a standard result under certain regularity conditions. As an example, the Log-AR(1) model (22) – (23) satisfies those conditions for $\beta \in (-1,1)$.

Proposition 1 *Let $(\Omega, \mathcal{F}, P)$ be a probability space on which the asset return process Y, volatility process h and parameter θ defined. If the MLE $\hat{\theta}(x^n)$ is consistent, then for every measurable set $A \subset \Theta$,*

$$\lim_{n\to\infty} P(A \mid Y^n = y^n) = I_A(\theta), \quad a.s. \quad under \ P. \tag{33}$$

Proof This proposition follows from Theorem 7.78 in Schervish (1995), provided we can show that

$$\xi_n = \xi_n(Y^n) = \int \hat{\theta}(x^n) \ f_{h^n}(h^n|Y^n) \ dh^n \tag{34}$$

is a consistent estimator for θ.

Since (34) implies $|\xi_n - \theta| \leq \int |\hat{\theta}(x^n) - \theta| \ f_{h^n}(h^n|Y^n) \ dh^n$, we conclude that ξ_n converges to θ in P-probability as $n \to \infty$. Therefore, Proposition 1 holds. *QED*.

An intuitive interpretation of this proposition is that the posterior distribution of θ will eventually concentrate near the true parameter value, with probability one under the joint distribution of the data and the parameter. Note that the prior $\pi(\theta)$ on Θ is not explicitly mentioned, which may or may not be supported in a neighborhood around the true parameter value.

Corollary 1 *Assume the same conditions as in the above proposition, and for every $\delta > 0$, the neighborhood $U_0(\delta) = \{\theta : |\theta - \theta_0| < \delta\}$ of a fixed $\theta_0 \in \Theta$ satisfies $\pi(U_0(\delta)) > 0$. Then*

$$\lim_{n\to\infty} P(A \mid Y^n = y^n) = 1, \quad a.s. \quad under \ P. \tag{35}$$

For the posterior distribution of a parameter, the asymptotic normality is a more relevant and informative property for what we present in this paper, because it indicates that one should expect to see a nearly "bell-curve" shaped histogram when plotting values of the parameter generated from the MCMC iterations. Following Le Cam (1986)

(Section 12.4), Hartigan (1983) (Chapter 11) and Schervish (1995) (Section 7.4.2), we mention the following general regularity conditions with the augmented data x^n, under which the asymptotic normality for $p(\theta|x^n)$ holds. Again as an example, those conditions can be verified (a tedious but standard argument) for the extended Log-AR(1) model with correlated noise factors.

The asymptotic normality for $p(\theta|y^n)$ will follow from the mixture representation

$$p(\theta|y^n) = \int p(\theta|x^n) \, p(dh^n|y^n). \tag{36}$$

General regularity conditions

(c1) The parameter space Θ is an open set in $I\!\!R^k$.

(c2) For a fixed $\theta_0 \in \Theta$, the prior density for θ is positive and continuous at θ_0.

(c3) There is a neighborhood $U_0 \subset \Theta$ of θ_0 on which $\nabla^2 l_{x^n}(\theta)$ is continuous in θ with

P_{θ_0} probability one.

(c4) The largest eigenvalue of Σ_{x^n} goes to zero in P_{θ_0}-probability.

(c5) For small $\delta > 0$, let $U_0(\delta) \subset \Theta$ be the open ball of radius δ centered at θ_0, and λ_n

be the smallest eigenvalue of Σ_{x^n}. Then there exists $b(\delta) > 0$ such that

$$\lim_{n\to\infty} P_{\theta_0} \left(\sup_{\theta \in \, \Theta \backslash U_0(\delta)} \lambda_n \, [l_{x^n}(\theta) - l_{x^n}(\theta_0)] < -b(\delta) \right) = 1.$$

(c6) For every $\epsilon > 0$, there exists $\delta > 0$, such that

$$\lim_{n\to\infty} P_{\theta_0} \left(\sup_{\theta \in \, U_0(\delta), \, \|v\|=1} \left| 1 + v^\top \Sigma_{x^n}^{\frac{1}{2}} \, \nabla^2 l_{x^n}(\theta) \, \Sigma_{x^n}^{\frac{1}{2}} \, v \right| < \epsilon \right) = 1.$$

Note: Among these conditions, (c1) prevents the likelihood function from having its peak on the boundary of Θ; (c2) ensures that the prior does not miss the true parameter value; (c4) implies that the amount of information in the data about the parameter increases without bound; (c5) guarantees that the MLE is consistent and the likelihood function is eventually dominated by its values near θ_0; (c6) is a smoothness condition on the Fisher information matrix.

The study of asymptotic normality is focused on the mixture density:

$$f_{\zeta_n}(z|y^n) = \int \ f_{Z_n}(z|x^n) \ f_{h^n}(h^n|y^n) \ dh^n, \tag{37}$$

where

$$Z_n = \Sigma_{x^n}^{-\frac{1}{2}}[\theta - \hat{\theta}(x^n)], \tag{38}$$

and $\zeta_n(Y^n)$ has the density $f_{\zeta_n}(\cdot \ |y^n)$ on Θ. Furthermore, ζ_n is a normalized random variable (similar to Z_n). The asymptotic mean of ζ_n is a mixture of $\hat{\theta}(x^n)$, and the asymptotic variance of ζ_n can be represented via its first and second moments, both as mixtures of the corresponding asymptotic moments of Z_n.

The necessity of normalization given in (38) can be seen from Proposition 1 where the unnormalized posterior distribution of θ tends to a degenerate measure, but not a normal distribution.

Let $\phi_k(z)$ be the density of the standard k-variate normal distribution $\mathcal{N}_k(0, I_k)$. Then we have

Proposition 2 *Assume that the distribution of the augmented data in SV models satis-*

fies the conditions (c1) — (c6). Then for fixed θ_0, $\epsilon > 0$ and any compact set $B \subset I\!\!R^k$,

$$\lim_{n \to \infty} P_{\theta_0} \left(\sup_{z \in B} |f_{\zeta_n}(z|Y^n) - \phi_k(z)| > \epsilon \right) = 0. \tag{39}$$

Proof   Let B be an arbitrary compact subset of $I\!\!R^k$. For every $z \in B$,

$$|f_{\zeta_n}(z|Y^n) - \phi_k(z)|$$

$$\leq \int |f_{Z_n}(z|x^n) - \phi_k(z)| \ f_{h^n}(h^n|y^n) \ dh^n$$

$$\leq \int \sup_{z \in B} |f_{Z_n}(z|x^n) - \phi_k(z)| \ f_{h^n}(h^n|y^n) \ dh^n.$$

Therefore, it suffices to show that for every $\epsilon > 0$ and any compact subset B of $I\!\!R^k$,

$$\lim_{n \to \infty} P_{\theta_0} \left(\sup_{z \in B} |f_{Z_n}(z|x^n) - \phi_k(z)| > \epsilon \right) = 0. \tag{40}$$

This is a standard version of asymptotic normality for posterior densities based on the

augmented data x^n, which follows from the conditions (c1) — (c6). QED.

Appendix

Lemma A.1 *The discounted asset price $S_t^* = e^{-\int_0^t r_u du} S_t$ is a Q-martingale if and only the condition (4) holds.*

Proof It is a straightforward verification for the sufficiency of (4). To show that (4) is necessary, assume that $\{S_t^*\}$ is a Q-martingale. It follows from (5) and Itô's formula that

$$S_t^* = S_0 + \int_0^t \xi_u du + \int_0^t \sigma_u \sqrt{1 - \rho_u^2}\ d\widetilde{W}_u^S + \int_0^t \sigma_u \rho_u\ d\widetilde{W}_u^\sigma, \tag{41}$$

where

$$\xi_t = -r_t + \mu_t - \sigma_t \left(\lambda_t \sqrt{1 - \rho_t^2} + \nu_t \rho_t \right). \tag{42}$$

This representation implies that $\{\int_0^t \xi_u du,\ 0 \le t \le T\}$ is a continuous Q-martingale of bounded variation with value zero at $t = 0$. Therefore, following Karatzas and Shreve (1991), P35, Problem 5.12, we have

$$P\left(\int_0^t \xi_u du = 0,\ \forall\ 0 \le t \le T \right) = 1,$$

then working with the positive and negative parts of ξ_t, we have

$$P\left(\int_B (\xi_u)^+ du = \int_B (\xi_u)^- du \quad \text{on every Borel subset } B \text{ of } [0, T] \right) = 1,$$

which means there exists a set $A \subset [0, T]$ with Lebesgue measure zero, such that

$$P(\xi_t = 0,\ \forall\ t \in A^c) = 1. \qquad QED.$$

References

Aït-Sahalia, Y., Wang, Y. and Yared, F. (1998). Do option markets correctly price the probabilities of movement of the underlying asset? Preprint.

Andersen, T. and Sorensen, B. (1995). GMM estimation of a stochastic volatility model: a Monte Carlo study. *Journal of Business and Economic Statistics* (?)

Avellaneda, M., Friedman, C., Holmes, R. and Samperi, D. (1998). Calibrating volatility surfaces via relative-entropy minimization. *Applied Mathematical Finance* **?**.

Bakshi, G., Cao, C. and Chen, Z. (1997). Empirical performance of alternative option pricing models. *Journal of Finance* **52**, 2003-2049.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* **31**, 307-327.

Campbell, J., Lo, A. and MacKinlay, C. (1997). *The Econometrics of Financial Markets*. Princeton University Press.

Carter, C.K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika* **81**, 541-553.

Chen, Y. (1997). *Bayesian Time Series: Financial Models and Spectral Analysis*. Ph.D. thesis at the Institute of Statistics and Decision Sciences of Duke University.

Chernov, M. and Ghysels, E. (2000). What data should be used to price options? *Journal of Financial Economics* **56**, 407-456.

Chesney, M. and Scott, L.O. (1989). Pricing European options: a comparison of the modified Black-Scholes model and a random variance model. *Journal of Financial and Quantitative Analysis* **24**, 267-284.

Chib, S., Nardari, F. and Shephard, N. (1998). Markov chain Monte Carlo methods for generalized stochastic volatility models. Preprint.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* **50**, 987-1007.

Fouque, J.-P., Papanicolaou, G.C. and Sircar, K.R. (1999). Mean-reverting stochastic volatility. To appear in *International Journal of Theoretical and Applied Finance*.

Gallant, A.R., Hsieh, D.A. and Tauchen, G. (1991). On fitting recalcitrant series: the pound/dollar echange rate. 1974-83. *Nonparametric and Semiparametric Methods in Economics and Statistics*. Cambridge University Press.

Gallant, A.R., Hsieh, D.A. and Tauchen, G. (1997). Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics* **81**, 159-192.

Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science* **7**, 457-511.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1998). *Bayesian Data Analysis*. Chapman & Hall.

Hartigan, J.A. (1983). *Bayes Theory*. Springer-Verlag.

Harvey, A., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies* **61**, 247-264.

Heston, S.L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* **6**, 327-343.

Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance* **42**, 281-300.

Jacquier, E., Polson, N.G. and Rossi, P.E. (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics* **12**, 371-417.

Jacquier, E., Polson, N.G. and Rossi, P.E. (1999). Stochastic volatility: univariate and multivariate extensions. Preprint.

Jacquier, E. and Jarrow, R. (2000). Bayesian analysis of contingent claim model error. *Journal of Econometrics* **94**, 145-180.

Jones, C.S. (2000). The dynamics of stochastic volatility. Preprint.

Karatzas, I. and Shreve, S.E. (1991). *Brownian Motion and Stochastic Calculus* (2nd edition). Springer-Verlag.

Kim, S., Shephard, N. and Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* **65**, 361-393.

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag.

Melino, A. and Turnbull, S.M. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics.* **45**, 239-265.

Øksendal, B. (1995). *Stochastic Differential Equations* (4th edition). Springer-Verlag.

Pitt, M. and Shephard, N. (1998). Time averaging covariances: a factor stochastic volatility approach (with discussion). *Bayesian Statistics* **6** (edited by J. Bernardo et al.) Oxford University Press.

Rebonato, R. (1996). *Interest Rate Option Models.* John Wiley & Sons.

Renault, E. and Touzi, N. (1996). Option hedging and implied volatilities in a stochastic volatility model. *Mathematical Finance* **6**, 279-302.

Romano, M. and Touzi, N. (1997). Contingent claims and market completeness in a stochastic volatility model. *Mathematical Finance* **7**, 399-412.

Schervish, M.J. (1995). *Theory of Statistics.* Springer-Verlag.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. *Time Series Models with Econometrics, Finance and Other Applications* (edited by D.R. Cox et al.), 1-67. Chapman & Hall.

Shephard, N. and Pitt, M. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* **84**, 653-667.

Taylor, S. (1994). Modeling stochastic volatility: a review and comparative study. *Mathematical Finance* **4**, 183-204.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics* **22**, 1701-1786.

Table 1: Posterior means and standard deviations

	posterior mean	posterior st.d.
α (-0.8)	-0.8580	0.2474
β (0.9)	0.8928	0.0303
η (0.6)	0.6263	0.0789

Table 2: Comparison study with different β values

	mean($\beta|y$)	st.d.($\beta|y$)	mean($\eta|y$)	st.d.($\eta|y$)
$\beta(0.9)$ $\eta(0.6)$	0.8928	0.0303	0.6263	0.0789
$\beta(0.8)$ $\eta(0.6)$	0.7370	0.0761	0.6361	0.1042
$\beta(0.7)$ $\eta(0.6)$	0.4657	0.1428	0.7292	0.1511

Table 3: Comparison study with different η values

	mean($\beta|y$)	st.d.($\beta|y$)	mean($\eta|y$)	st.d.($\eta|y$)
$\beta(0.9)$ $\eta(0.6)$	0.8928	0.0303	0.6263	0.0789
$\beta(0.9)$ $\eta(0.2)$	0.8470	0.0941	0.2613	0.1249

Table 4: Comparison study with different sample sizes

Sample Size	mean($\beta|y$)	st.d.($\beta|y$)	mean($\eta|y$)	st.d.($\eta|y$)
T=100	0.7269	0.1188	0.8231	0.1743
T=200	0.8668	0.0795	0.5301	0.1834
T=300	0.8718	0.0414	0.7633	0.1045
T=500	0.8928	0.0303	0.6263	0.0789

Table 5: Posterior means and standard deviations

	posterior mean	posterior st.d.
α (-0.8)	-0.7768	0.1739
β (0.9)	0.9080	0.0205
η (0.6)	0.5864	0.0311

Table 6: Comparison between historical and hybrid volatilities

		historical volatility		hybrid volatility	
		mean	st.d.	mean	st.d.
T=50	α (-0.8)	-7.8022	5.3454	-1.2813	0.8823
	β (0.9)	0.1082	0.6100	0.8093	0.1355
	η (0.6)	0.7527	0.1522	0.7021	0.1005
T=100	α (-0.8)	-2.1566	0.9341	-1.6474	0.5422
	β (0.9)	0.7269	0.1188	0.7754	0.0729
	η (0.6)	0.8231	0.1743	0.6794	0.0805
T=200	α (-0.8)	-1.0918	0.6513	-0.8203	0.3628
	β (0.9)	0.8668	0.0795	0.8982	0.0465
	η (0.6)	0.5301	0.1834	0.5217	0.0752
T=500	α (-0.8)	-8580	0.2474	-0.7768	0.1739
	β (0.9)	0.8928	0.0303	0.9080	0.0205
	η (0.6)	0.6263	0.0789	0.5864	0.0311

Table 7: Volatlity smile in stochastic volatility model

Strike Price	Option Price	BS Implied Vol.
100	24.4	0.0287
105	20.6	0.0284
110	17.1	0.0277
115	14.0	0.0276
120	11.3	0.0273
125	9.1	0.0274
130	7.5	0.0282
135	5.8	0.0278
140	4.6	0.0279
145	3.9	0.0292
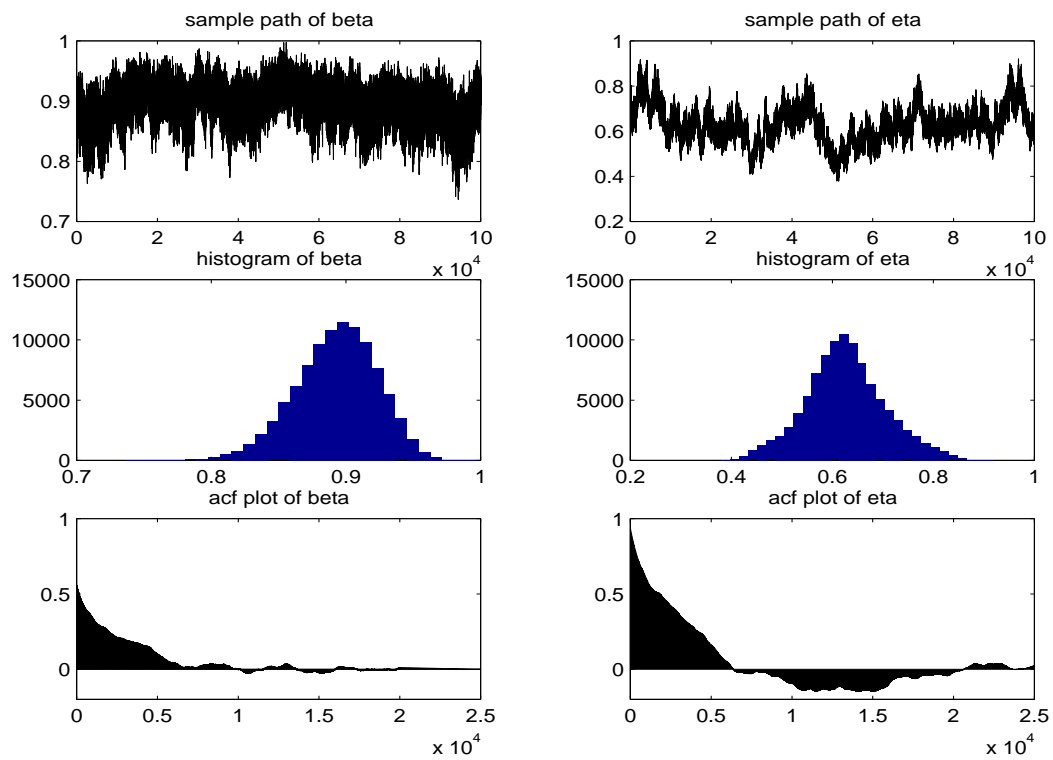
Figure 1: Trajectory plots, histograms and acf plots of β and η
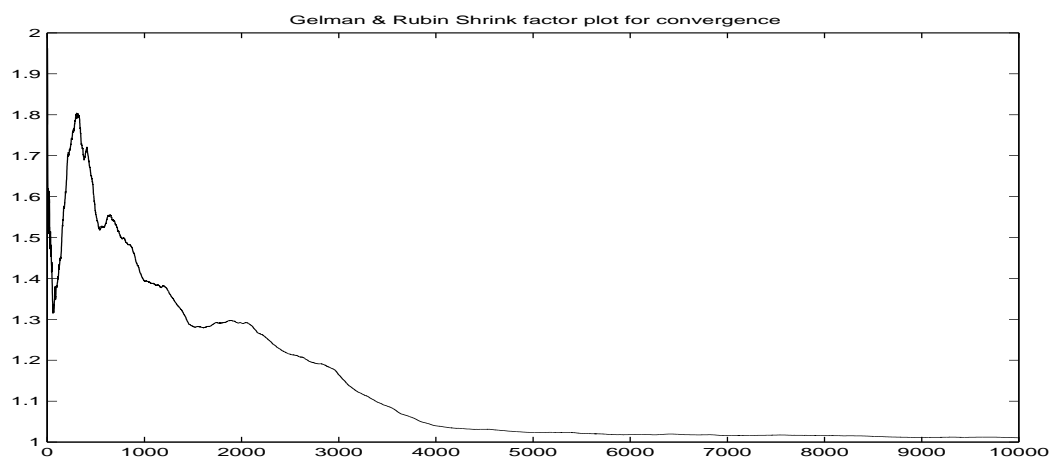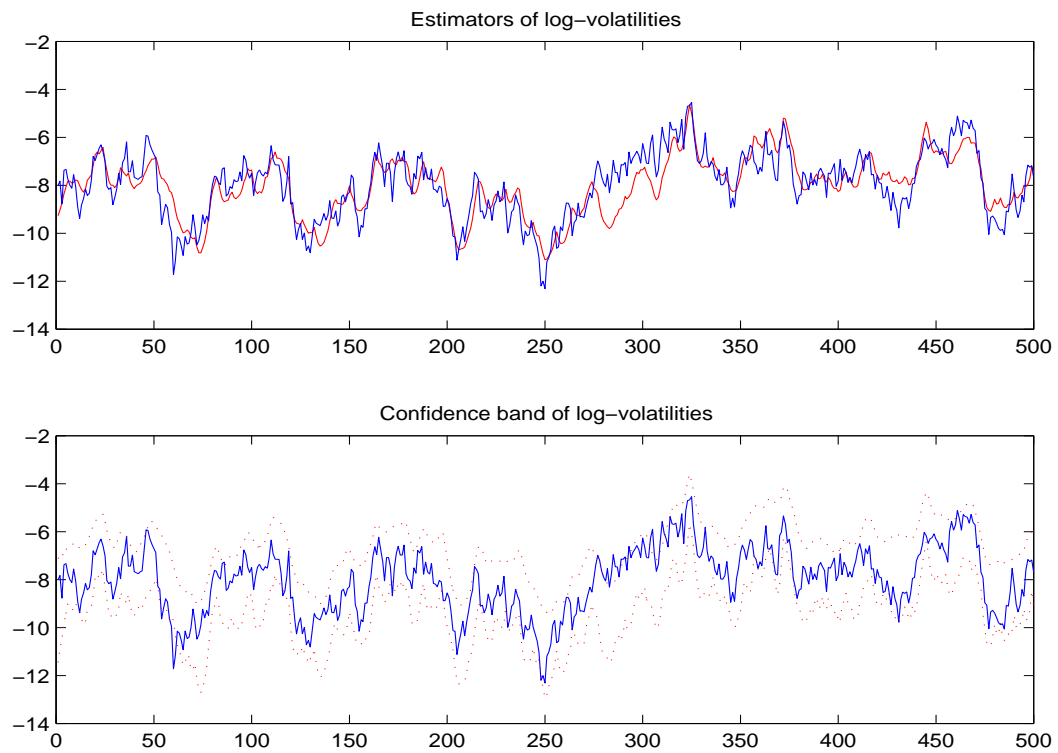

Figure 2: Test of MCMC convergence plot

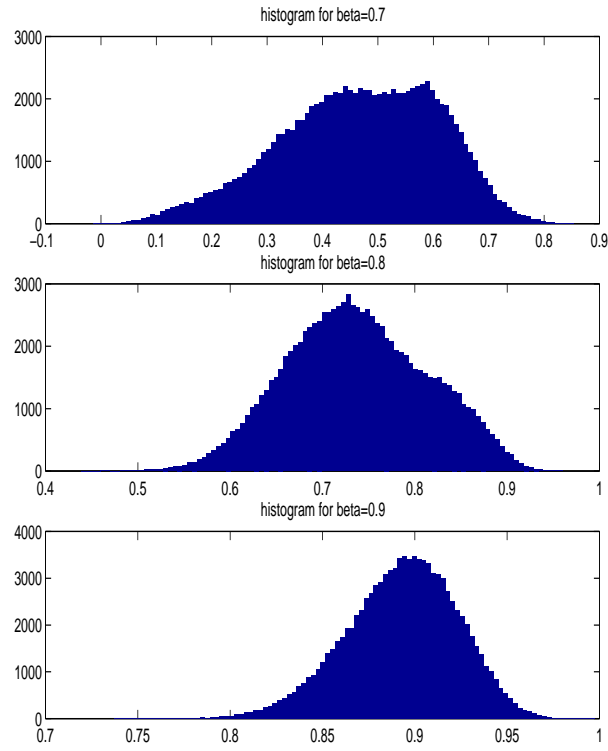Figure 3: Volatility summaries as a byproduct of MCMC

Figure 4: Histogram plots of β with different true β values
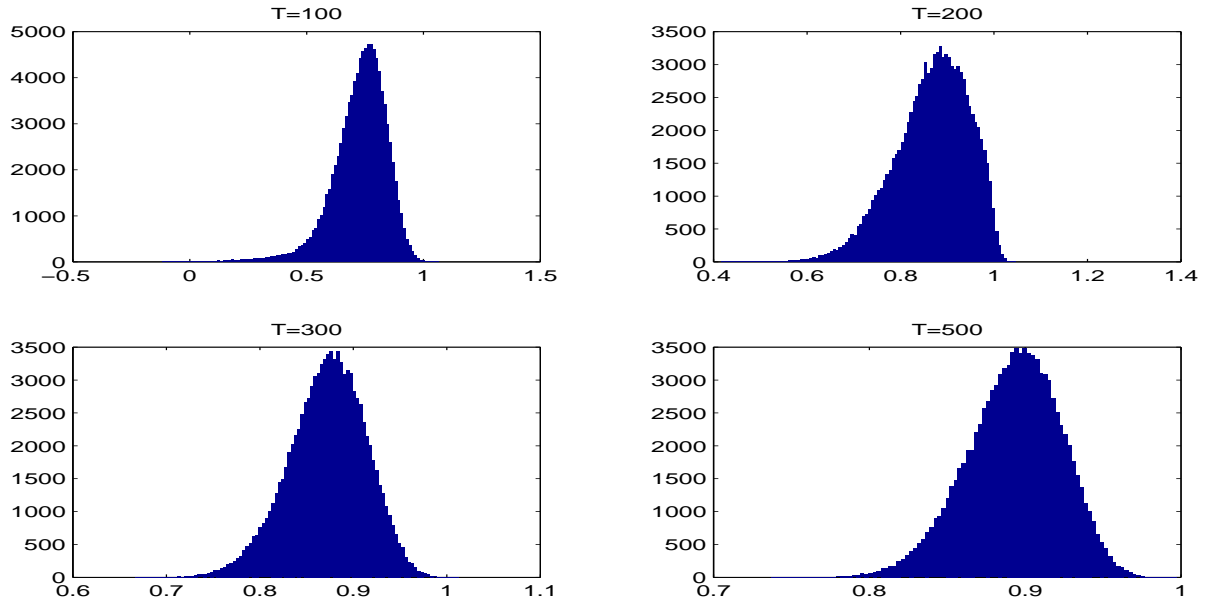

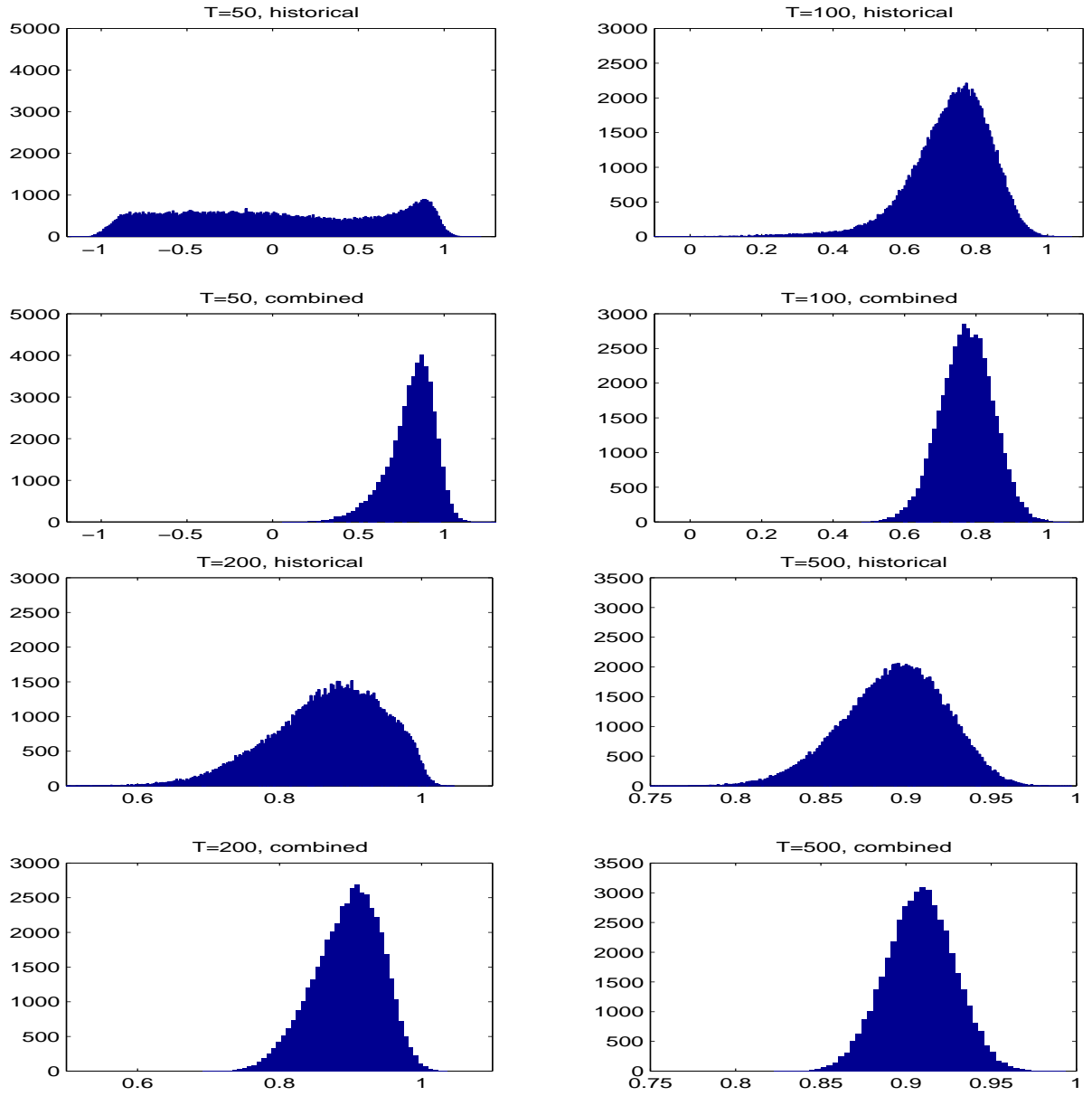Figure 5: Histogram plots of β with different sample sizes

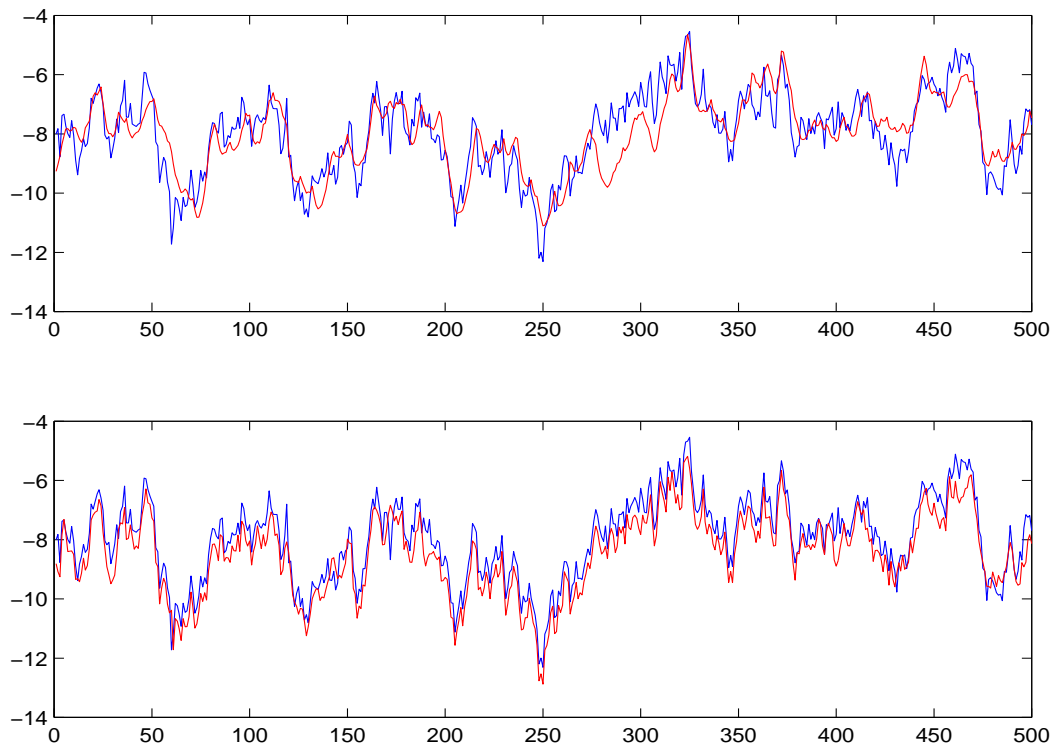Figure 6: Comparison of histogram plots of β
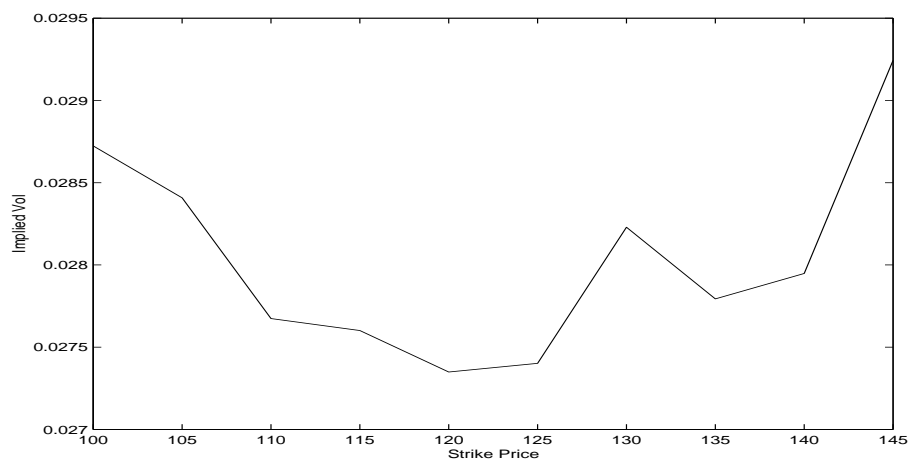
Figure 7: Comparison of volatility estimation

Figure 8: Volatility smile in stochastic volatility model