RESEARCH PAPER
quant.iop.org

QUANTITATIVE FINANCE VOLUME 4 (2004) 158–169
INSTITUTE OF PHYSICS PUBLISHING

Estimation of hyperbolic diffusion using the Markov chain Monte Carlo method

Y K Tse[1], Xibin Zhang[2] and Jun Yu[3]

[1] School of Economics and Social Sciences, Singapore Management University, 259756, Singapore
[2] Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria 3800, Australia
[3] Department of Economics, University of Auckland, Auckland, New Zealand

E-mail: yktse@smu.edu.sg, bill.zhang@buseco.monash.edu.au and j.yu@auckland.ac.nz

Received 29 September 2003
Published 3 December 2003
Online at stacks.iop.org/Quant/4/158 (DOI: 10.1088/1469-7688/4/2/005)

Abstract

In this paper we propose a Bayesian method to estimate the hyperbolic diffusion model. The approach is based on the Markov chain Monte Carlo (MCMC) method with the likelihood of the discretized process as the approximate posterior likelihood. We demonstrate that the MCMC method provides a useful tool in analysing hyperbolic diffusions. In particular, quantities of posterior distributions obtained from the MCMC outputs can be used for statistical inference. The MCMC method based on the Milstein scheme is found to perform well with good mixing properties, while the Euler scheme is unsatisfactory. Our simulation study shows that the hyperbolic diffusion exhibits many of the stylized facts about asset returns documented in the discrete-time financial econometrics literature, such as the Taylor effect, a slowly declining autocorrelation function of the squared returns, and thick tails.

1. Introduction

In the finance literature the geometric Brownian motion has been used as a classical model to describe stock price movements. Though useful as a benchmark model in option pricing and other theories, the geometric Brownian motion is irreconcilable with many known statistical regularities of stock returns, such as excess kurtosis, clustering of volatility and long memory. To this effect, other processes have been suggested, such as jump diffusions (Kou 2002), stochastic volatility (SV) models (Heston 1993), SV plus jumps (Eraker *et al* 2003), and time-changed Levy process (Carr and Wu 2003). As a nonlinear diffusion process, the hyperbolic diffusion model proposed by Bibby and Sorensen (1997) has received some attention (see, e.g., Rydberg 1999). Bibby and Sorensen (1997) demonstrated some success in fitting the stationary distribution of the hyperbolic diffusion to some stock price data, and provided the theory in applying the hyperbolic diffusion to option pricing.

Although the stationary distribution of the hyperbolic diffusion process follows the hyperbolic distribution and hence has a closed-form expression, the transition density has no closed-form solution. Due to lack of knowledge of the transition density, econometric estimation of the model using the *exact* likelihood approach is intractable, though an approximate likelihood method based on discretization may be adopted. To circumvent this difficulty Bibby and Sorensen (1997) estimated the hyperbolic diffusion using the martingale

1469-7688/04/020158+12$30.00 PII: S1469-7688(04)69558-3

estimating function method. However, although the estimator based on martingale estimating functions is consistent and asymptotically normally distributed, it is inefficient in general. Furthermore, computation of the standard errors of the resulting estimates is difficult and requires techniques such as parametric bootstrapping.

In this paper we propose to use the Markov chain Monte Carlo (MCMC) method to estimate the parameters of the hyperbolic diffusion with the discretized likelihood of the diffusion process as an approximate posterior. Like the maximum likelihood (ML) approach in the classical framework, the MCMC method offers a full likelihood-based inference based on Bayesian analysis. In the case of the hyperbolic diffusion, the discretized approximate ML approach is found to encounter difficulties in numerical convergence. The MCMC method, however, provides a general mechanism to sample the parameter vector from its posterior distribution, and hence avoids the need for numerical optimization and enables exact finite-sample inferences via Monte Carlo methods.

In the financial econometrics literature a number of stylized facts have been well documented in describing the statistical properties of equity return series. Several models in the discrete-time domain have been found to be able to generate time series with such stylized properties. In the continuous-time domain, however, the success in this aspect has been much weaker[4]. On the other hand, while most empirical works rely on discrete-time models due to their simplicity in estimation, theories in option pricing are usually based on continuous-time models. The hyperbolic diffusion is a promising continuous-time model that describes empirically equity price data and can be applied to option pricing. Empirical illustrations reported by Rydberg (1999) show that a member of the generalized hyperbolic diffusion can induce long-memory features in the squared return[5]. In this paper we report the ability of the hyperbolic diffusion in reproducing other stylized facts documented in the financial econometrics literature.

This paper is organized as follows. Section 2 reviews the hyperbolic diffusion model and its properties. We discuss how the Euler and Milstein schemes can be used to discretize the model, and thus to provide approximations to the posterior likelihood. Some stylized facts about equity return series are summarized and related to the hyperbolic diffusion. Section 3 describes the MCMC method. In section 4 we fit the model to three stock market indexes over a decade of daily data using the MCMC method based on both the Milstein and Euler schemes. Statistical inference is then made via the posterior quantities. In section 5 we examine the statistical properties of sample paths generated by the hyperbolic diffusion. We find that many of the stylized facts for stock returns in the empirical finance literature documented by Ryden *et al* (1998) are satisfied. Section 6 concludes.

[4] A notable exception is the time-changed Levy process proposed recently by Carr and Wu (2003). These authors, however, did not provide any empirical analysis of their model. Eraker *et al* (2003) proposed a diffusion model with jumps in both the return and the volatility, but they did not examine the statistical properties of their process.

[5] Rydberg (1999) considered the normal inverse Gaussian diffusion, which differs from the hyperbolic diffusion in the form of the stationary density.

2. Hyperbolic diffusion and some stylized facts of stock returns

Consider the following continuous-time parametric diffusion:

$$\mathrm{d}X_t = \mu(X_t, \theta)\,\mathrm{d}t + \sigma(X_t, \theta)\,\mathrm{d}W_t, \tag{1}$$

where X_t is a state variable, W_t is a standard Brownian motion defined on the probability space $(\Omega, \Im^{\mathrm{B}}, (\Im_t^{\mathrm{B}})_{t\geqslant 0}, P)$, $\mu(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ are known functions, and θ is a vector of unknown parameters.

Many empirical studies have shown that asset returns are not normally distributed. Barndorff-Nielsen (1978) suggested using the hyperbolic distribution to describe unconditional asset returns. The density of the hyperbolic distribution is proportional to $1/b^2(x)$, with

$$b(x) = \exp\{\tfrac{1}{2}[\alpha\sqrt{\delta^2 + (x-\mu)^2} - \beta(x-\mu)]\}, \tag{2}$$

where α, β, δ and μ are the parameters of the distribution satisfying $\alpha > |\beta| \geqslant 0$ and $\delta > 0$. It is noted that δ is the scale parameter, μ is the location parameter, β determines the symmetry (the distribution is symmetrical about μ if $\beta = 0$) and α determines the steepness of the distribution.

We assume that the stock price S_t depends on the state variable X_t as follows:

$$S_t = \exp(X_t + \kappa t), \tag{3}$$

where κ is the (constant) drift rate. Following Bibby and Sorensen (1997) we consider the following hyperbolic diffusion process to describe the movement of stock prices[6]:

$$\mathrm{d}S_t = S_t\{[\kappa + \tfrac{1}{2}\sigma^2 b^2(\ln S_t - \kappa t)]\,\mathrm{d}t + \sigma b(\ln S_t - \kappa t)\,\mathrm{d}W_t\}. \tag{4}$$

Bibby and Sorensen (1997) obtained some interesting statistical properties of the process S_t. For instance, they showed that the marginal distribution of $\ln S_t$ is hyperbolic and hence $\ln S_t$ is approximately hyperbolically distributed after a sufficiently long time period. Also, the distribution of increments over short intervals has thick tails while an increment over a long interval follows a distribution that is close to being hyperbolic.

To derive the dynamic properties of stock returns, we apply Ito's lemma to obtain

$$\mathrm{d}X_t = \sigma b(X_t)\,\mathrm{d}W_t, \tag{5}$$

which represents a diffusion process with no drift. As $\mathrm{d}W_t$ are uncorrelated over nonoverlapping intervals, increments of the log-prices (i.e. the continuously compounded rates of return) are serially uncorrelated. Similar to the SV and autoregressive conditional heteroscedasticity (ARCH) models, the squared increments of the log-prices are generally serially correlated. In other words, return series of hyperbolic diffusion are likely to exhibit volatility clustering as demonstrated by SV and

[6] Note that μ in equation (2) and σ in equation (4) are parameters of the diffusion. They should not be confused with $\mu(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$, which are known functions of the drift and diffusion terms, respectively.

ARCH models, which have been found to be successful in describing many stylized facts of equity return series.

To understand why a hyperbolic diffusion generates volatility clustering and long memory properties, we apply the Euler approximation to the diffusion model for the log-price (i.e. equation (1)) and obtain

$$Y_t \approx \sigma \exp\left[\frac{1}{2}\left\{\alpha\sqrt{\delta^2 + \left(\sum_{i=1}^{\infty} Y_{t-i} - \mu\right)^2} - \beta\left(\sum_{i=1}^{\infty} Y_{t-i} - \mu\right)\right\}\right] e_t, \tag{6}$$

where $Y_t = \ln S_{t+\Delta t} - \ln S_t$ denotes the return and $e_t \sim$ i.i.d. $N(0, \Delta t)$. Equivalently this equation can be rewritten as

$$Y_t \approx \sigma \exp\{\tfrac{1}{2}h_t\}e_t$$

$$h_t = \alpha\sqrt{\delta^2 + \left(\sum_{i=1}^{\infty} Y_{t-i} - \mu\right)^2} - \beta\left(\sum_{i=1}^{\infty} Y_{t-i} - \mu\right).$$

Comparing the above specification with the well-known ARCH(∞) model (Engle 1982),

$$Y_t = \sigma_t e_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{\infty} \alpha_i Y_{t-i}^2,$$

it can be seen that the hyperbolic diffusion model can be regarded as a special case of the following nonlinear ARCH(∞) model:

$$Y_t = \sigma \exp\{\tfrac{1}{2}h_t\}e_t$$

$$h_t = f(Y_{t-1}, Y_{t-2}, \ldots).$$

This suggests that the hyperbolic diffusion model may generate return series that exhibit ARCH effects. Furthermore, the nonlinear relationship between h_t and Y_{t-i} may cause long-memory properties in the absolute return as well as the squared return.

Before we discuss the MCMC estimation methods for the hyperbolic diffusion we summarize here some stylized facts for equity return series in the empirical finance literature, which may be used as benchmarks for empirical equity price processes. Let r_t denote the return of a stock. Ryden *et al* (1998) summarized the following dynamic properties of r_t found in many empirical studies:

(1) r_t are not autocorrelated, except possibly at lag one.
(2) The autocorrelation functions (ACFs) of $|r_t|$ and r_t^2 decay slowly. The decay is much slower than the exponential rate of the ACF of a stationary ARMA process.
(3) $\mathrm{corr}(|r_t|, |r_{t-k}|) > \mathrm{corr}(|r_t|^\phi, |r_{t-k}|^\phi)$, $\phi \neq 1$. The autocorrelations of powers of absolute return are highest at power one. This is called the Taylor effect.

In additional to the above dynamic properties, the following are two well known static properties:

(1) Returns often show strong evidence that the marginal distribution has thick tails.
(2) Returns often show weak evidence that the marginal distribution is skewed.

3. Discretization of hyperbolic diffusion

Although the unconditional distribution of the hyperbolic diffusion process is hyperbolic, the transition density is unknown. Therefore, the exact ML method is difficult to implement. Bibby and Sorensen (1997) suggested using the martingale estimating function approach of Bibby and Sorensen (1995) to estimate the diffusion models. This approach, however, requires knowledge of the conditional expectation and conditional variance of the underlying diffusion which are known only for very simple models, such as those with a linear drift in the state variable. Hence, although the martingale estimating function method provides estimates that are consistent and asymptotically normal, implementation of the method is difficult in practice.

We propose to use the MCMC method to estimate the hyperbolic diffusion after discretizing the model. In the next section we shall discuss the MCMC method. In this section, we outline the Euler and Milstein schemes for the discretization of the hyperbolic diffusion model. The discretized schemes provide approximations to the likelihood function, which is in turn used to approximate the posterior to facilitate Bayesian analysis. It is well known that the Milstein scheme provides an approximation with improved accuracy over the Euler scheme on approximating the underlying diffusion (Milstein 1978, Kloeden and Platen 1992). As a consequence, it is expected that the likelihood and posterior calculated from the Milstein scheme would provide better approximations to the true counterparts than those from the Euler scheme. Indeed, Elerian (1998) compared the performance of the ML method based on these two schemes in the context of a univariate CIR model (Cox *et al* 1985) and found that the Milstein scheme offers improvements over the Euler scheme.

The Euler scheme approximates a general diffusion process such as equation (1) by the following expansion:

$$X_{t+\Delta t} = X_t + \mu(X_t, \theta)\Delta t + \sigma(X_t, \theta)\Delta W_t,$$

where $\Delta W_t = \varepsilon_t\sqrt{\Delta t}$ with $\varepsilon_t \sim$ i.i.d. $N(0, 1)$. Assuming constant priors for all the parameters, and given $n+1$ observations of $\mathbf{x} = \{x_t : t = 0, 1, \ldots, n\}$, the logarithmic likelihood of θ upon dropping the constant term, is

$$\log p_E(\theta|\mathbf{x}) = -\tfrac{1}{2}\sum_{t=1}^{n} \log(\sigma(x_t, \theta)^2\Delta t) - \frac{1}{2}\sum_{t=1}^{n} \frac{(x_t - x_{t-1} - \mu(x_t, \theta)\Delta t)^2}{\sigma(x_t, \theta)^2\Delta t}, \tag{7}$$

which can be used directly as an approximate logarithmic likelihood for ML estimation or as an approximate logarithmic posterior for the MCMC algorithm. Hereafter, we refer to $p_E(\theta|\mathbf{x})$ defined in (7) as the Euler likelihood.

Taking a higher-order term in the Taylor expansion, the Milstein approach approximates a general diffusion process by the following equation:

$$X_{t+\Delta t} = X_t + \mu(X_t, \theta)\Delta t + \sigma(X_t, \theta)\Delta W_t + \frac{1}{2}\sigma(X_t, \theta)\frac{\partial\sigma(X_t, \theta)}{\partial X_t}[(\Delta W_t)^2 - \Delta t], \tag{8}$$

which can be rewritten as

$$X_{t+\Delta t} - X_t - \mu(X_t,\theta)\Delta t + g(X_t,\theta)\Delta t$$
$$= \sigma(X_t,\theta)\sqrt{\Delta t}\,\varepsilon + g(X_t,\theta)\Delta t\varepsilon^2, \tag{9}$$

where $g(X_t,\theta) = \frac{1}{2}\sigma(X_t,\theta)(\partial\sigma(X_t,\theta)/\partial X_t)$. Let

$$a = \sigma(X_t,\theta)\sqrt{\Delta t}, \qquad b = g(X_t,\theta)\Delta t, \tag{10}$$

then equation (8) can be represented by

$$Y = a\varepsilon + b\varepsilon^2 = b\left[\left(\varepsilon + \frac{a}{2b}\right)^2 - \frac{a^2}{4b^2}\right], \tag{11}$$

where $Y = X_{t+\Delta t} - X_t - \mu(X_t,\theta)\Delta t + g(X_t,\theta)\Delta t$.

The normality assumption implies that $(\varepsilon + \frac{a}{2b})^2$ (denoted by, say, Z) follows a noncentral χ^2 distribution with one degree of freedom and noncentrality parameter $\lambda = a^2/(4b^2)$. Elerian (1998) showed that the density of Z is given by

$$f(z) = \frac{1}{2}\exp\left\{-\frac{\lambda + z}{2}\right\}\left(\frac{z}{\lambda}\right)^{-1/4} I_{-1/2}\big(\sqrt{\lambda z}\big), \tag{12}$$

where

$$I_{-1/2}(w) = \sqrt{\frac{2}{w}}\sum_{j=0}^{\infty}\frac{(w/2)^{2j}}{j!\Gamma(j+1/2)} = \sqrt{\frac{2}{\pi w}}\cosh(w),$$

with $\cosh(w) = \frac{1}{2}\{\exp(w) + \exp(-w)\}$ being the hyperbolic cosine function. Hence the density of Y is

$$f^*(y) = \frac{1}{b}f\left(\frac{y}{b} + \frac{a^2}{4b^2}\right) \tag{13}$$

and the logarithmic likelihood upon dropping the constant is given by

$$\log p_{\mathrm{M}}(\theta|\mathbf{x}) = \sum_{t=1}^{n}\left[\log\left\{f\left(\frac{y_t}{b} + \frac{a^2}{4b^2}\right)\right\} - \log(b)\right], \tag{14}$$

where $y_t = x_t - x_{t-1} - \mu(x_{t-1},\theta)\Delta t + g(x_{t-1},\theta)\Delta t$. Again, the above quantity can be used either directly as an improved approximate logarithmic likelihood for ML estimation or as an improved approximate logarithmic posterior for the purpose of MCMC simulation. Hereafter, we refer to $p_{\mathrm{M}}(\theta|\mathbf{x})$ defined in (14) as the Milstein likelihood.

4. Estimating hyperbolic diffusion via MCMC

4.1. MCMC

The MCMC strategy has proved useful in many statistical applications, and has many advantages compared to traditional independent sampling methods. Geweke (1999) provided a survey of the fundamental principles of subjective Bayesian inference in econometrics and the implementation of these principles using posterior simulation methods, emphasizing the importance of simulation methods and describing the implementation of MCMC simulation for Bayesian inference. Gilks *et al* (1996) presented a collection of papers on the application of MCMC algorithms. In econometrics and finance many successful applications of the MCMC method can be found (e.g., Eraker 2001, Elerian *et al* 2002). We refer readers to Chib (2001) and Johannes and Polson (2003) for recent surveys on the applications of MCMC in econometrics and finance.

Bayesian inference concerning a parameter vector θ conditional on data $\mathbf{x}$ is made via the posterior density $p(\theta|\mathbf{x})$. By the Bayes theorem, the posterior takes the form

$$\pi(\theta|\mathbf{x}) = cp(\mathbf{x}|\theta)\pi(\theta), \tag{15}$$

where c is a normalizing constant, $p(\mathbf{x}|\theta)$ is the likelihood of $\mathbf{x}$ conditional upon θ and $\pi(\theta)$ is the prior density of θ. The Bayesian approach requires that statistical inference be based on the posterior. Dealing with the posterior, however, is often analytically intractable. Nonetheless, if we can sample the parameter vector from the posterior, statistical inference about the parameter vector can be made using the usual Monte Carlo approach. The MCMC method aims to provide a general mechanism to sample the parameter vector from its posterior density. While simulating directly from the posterior distribution is typically very difficult, the MCMC method sets up a Markov chain so that its stationary distribution is the same as the posterior density. When the Markov chain converges, the simulated values may be regarded as a sample obtained from the posterior.

There are two broad categories of algorithms for implementing MCMC, which are, respectively, the Gibbs sampler and the Metropolis–Hastings algorithm. Let the current state be denoted as $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$, and assume that the full conditional densities of θ_i are available. The Gibbs sampler generates the next state θ', in which each component is generated from a sequence of conditional densities. The Metropolis–Hastings algorithm generates a candidate θ' from a proposal density denoted by $q(\cdot|\theta)$. The proposal density should satisfy certain properties, such as the reversibility condition discussed in Chib and Greenberg (1995) and Gilks *et al* (1996), among many others. The candidate is then accepted with probability $T(\theta,\theta')$, which is defined by

$$T(\theta,\theta') = \min\left\{1, \frac{\pi(\theta'|\mathbf{x})q(\theta|\theta')}{\pi(\theta|\mathbf{x})q(\theta'|\theta)}\right\}. \tag{16}$$

If the candidate is accepted, the next state is set to θ'. Otherwise, the chain does not move. Robert and Casella (1999, chapter 7) showed that the Gibbs sampler is equivalent to a composition of p Metropolis–Hastings algorithms with acceptance probabilities uniformly equal to 1. Robert and Casella (1999) presented detailed discussions on the use of the Metropolis–Hastings algorithm and the Gibbs sampler.

As the full conditional density is often difficult to derive, the Metropolis–Hastings algorithm is generally adopted in complex problems. In this paper we use the Metropolis–Hastings algorithm for its simplicity. In what follows we briefly describe the procedure of the algorithm.

Step 1. Given the current state $\theta^{(i)}$, generate a candidate θ' from the proposal density $q(\cdot|\theta^{(i)})$.

Step 2. Calculate the acceptance probability $T(\theta^{(i)}, \theta')$ according to (16).

Step 3. Accept the proposal with probability $T(\theta^{(i)}, \theta')$ and set $\theta^{(i+1)} = \theta'$. Otherwise, reject the candidate and set $\theta^{(i+1)} = \theta^{(i)}$.

Step 4. Repeat the previous steps to obtain a chain $\{\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \ldots\}$, where $\theta^{(0)}$ denotes the initial state of θ. Discard the burn-in values (up to $\theta^{(d)}$, say) obtained whilst the chain converges in distribution to the joint posterior. Then the remaining values, $\{\theta^{(d+1)}, \theta^{(d+2)}, \ldots\}$, are a correlated chain simulated from $\pi(\theta|\mathbf{x})$, and have the same stationary transition density as $\pi(\theta|\mathbf{x})$.

Two important points should be noted. First, the calculation of $T(\theta^{(i)}, \theta')$ does not require knowledge of the normalizing constant in the posterior function. Second, if the proposal density is symmetric, that is $q(x|y) = q(y|x)$, then the acceptance probability reduces to $\pi(\theta'|\mathbf{x})/\pi(\theta^{(i)}|\mathbf{x})$. Moreover, if $q(y|x)$ is a function of $|y - x|$, the resulting algorithm is called the random-walk Metropolis–Hastings algorithm, which has been widely used in practice due to its simplicity.

4.2. Empirical results

In this section we apply the random-walk Metropolis–Hastings algorithm to the discretized diffusion processes and present empirical results based on some real data sets. The data series considered are the MSCI World index, the MSCI Europe index and the NYSE index. The series consist of weekly observations from 1 January 1990 to 31 December 2000.

As argued before, the transition density of hyperbolic diffusions does not have closed-form, making the direct ML approach difficult. However, the transition density of discretized models under both schemes has an analytic expression, which, in theory, can be used to obtain the approximate ML estimates. Before we carried out the Bayesian MCMC analysis, we implemented the approximate ML estimation but found that numerical optimizations rarely converged. This experience indicates that the likelihood function of the discretized models is not well behaved. As a conditional simulation method, MCMC avoids any numerical difficulties associated with numerical optimizations. We now describe the details of the implementation of the MCMC method for the estimation of the hyperbolic diffusion.

4.2.1. Empirical results under the Milstein likelihood. Assume that the priors of the parameters are given by $\kappa \sim N(0, 10)$, $\alpha \sim \Gamma(1, 20)$, $\delta^2 \sim \Gamma(0.05, 20)$, $\mu \sim N(5, 10)$, $\beta \sim U(-\alpha, \alpha)$ and $\sigma^2 \sim IG(5, 0.05)$, where U, Γ, IG refer to the uniform, gamma and inverted gamma densities, respectively. These priors are very flat and nearly noninformative. The joint prior of all the parameters, denoted as $\pi(\theta)$, is the product of these marginal priors. Based on the Milstein likelihood $p_{\mathrm{M}}(\theta|\mathbf{x})$, we can obtain the joint posterior

$$\pi(\theta|x) \propto \pi(\theta)p_{\mathrm{M}}(\theta|x).$$

In the implementation of the random-walk Metropolis–Hastings algorithm, the proposal density is uniform on $[-0.5, 0.5]$, and the parameter vector θ is updated as follows:

$$\theta' = \theta + \tau\varepsilon,$$

where θ' is the proposal for θ, ε is a vector of random numbers drawn from the uniform density on $[-0.5, 0.5]$, and τ is a tuning parameter which is chosen so that the acceptance rate is between 20% and 30%. In addition, τ may be either a scalar- or vector-constant. Generally speaking, if the parameters are of weak correlation and their values are of the same scale, τ can be a scalar constant. Otherwise, τ should be a constant vector, so that each parameter is assigned a specific tuning parameter.

4.2.2. Convergence checking. In the implementation of the MCMC algorithm, the sampled path, denoted by $\{\theta^{(i)} : i = 1, 2, \ldots, N\}$, forms a Markov chain whose stationary density is the posterior $\pi(\theta|\mathbf{x})$, and the output is summarized in terms of the ergodic averages in the form of

$$\bar{f}_N = \frac{1}{N}\sum_{i=1}^{N} f(\theta^{[i]}), \tag{17}$$

where $f(\cdot)$ is a real-valued function to be estimated. Roberts (1996) pointed out that most of the Markov chains produced in MCMC converge geometrically to the stationary distribution $\pi(\theta|\mathbf{x})$, and one of the most important consequences of the geometric convergence is that the central limit theorem of ergodic averages is invoked, i.e.

$$\sqrt{N}(\bar{f}_N - E_\pi[f(\theta)]) \xrightarrow{D} N(0, \sigma_f^2), \tag{18}$$

where $E_\pi[\cdot]$ denotes the expectation operator under $\pi(\theta|\mathbf{x})$, and the convergence is in distribution. To assess the accuracy of the ergodic average as an estimate of $E_\pi[f(\theta)]$, it is essential to estimate σ_f^2. One of the most commonly used methods for estimating σ_f^2 is the batch mean, which is discussed extensively in Roberts (1996).

To estimate σ_f^2 using the batch mean, the MCMC algorithm is run for $N = m \times n$ iterations, where n is sufficiently large so that

$$y_k = \frac{1}{n}\sum_{i=(k-1)n+1}^{kn} f(\theta^{[i]}), \tag{19}$$

for $k = 1, 2, \ldots, m$, are approximately independently distributed as $N(E_\pi[f(\theta)], \sigma_f^2/n)$. Therefore σ_f^2 can be estimated by

$$\hat{\sigma}_f^2 = \frac{n}{m-1}\sum_{k=1}^{m}(y_k - \bar{f}_N)^2, \tag{20}$$

where $\bar{f}_N$ is defined in equation (17). Thus, the standard error of $\bar{f}_N$ can be estimated by $\sqrt{\hat{\sigma}_f^2/N}$, which is called the batch-mean standard error or the Monte Carlo standard error, and is commonly used for checking the mixing performance.

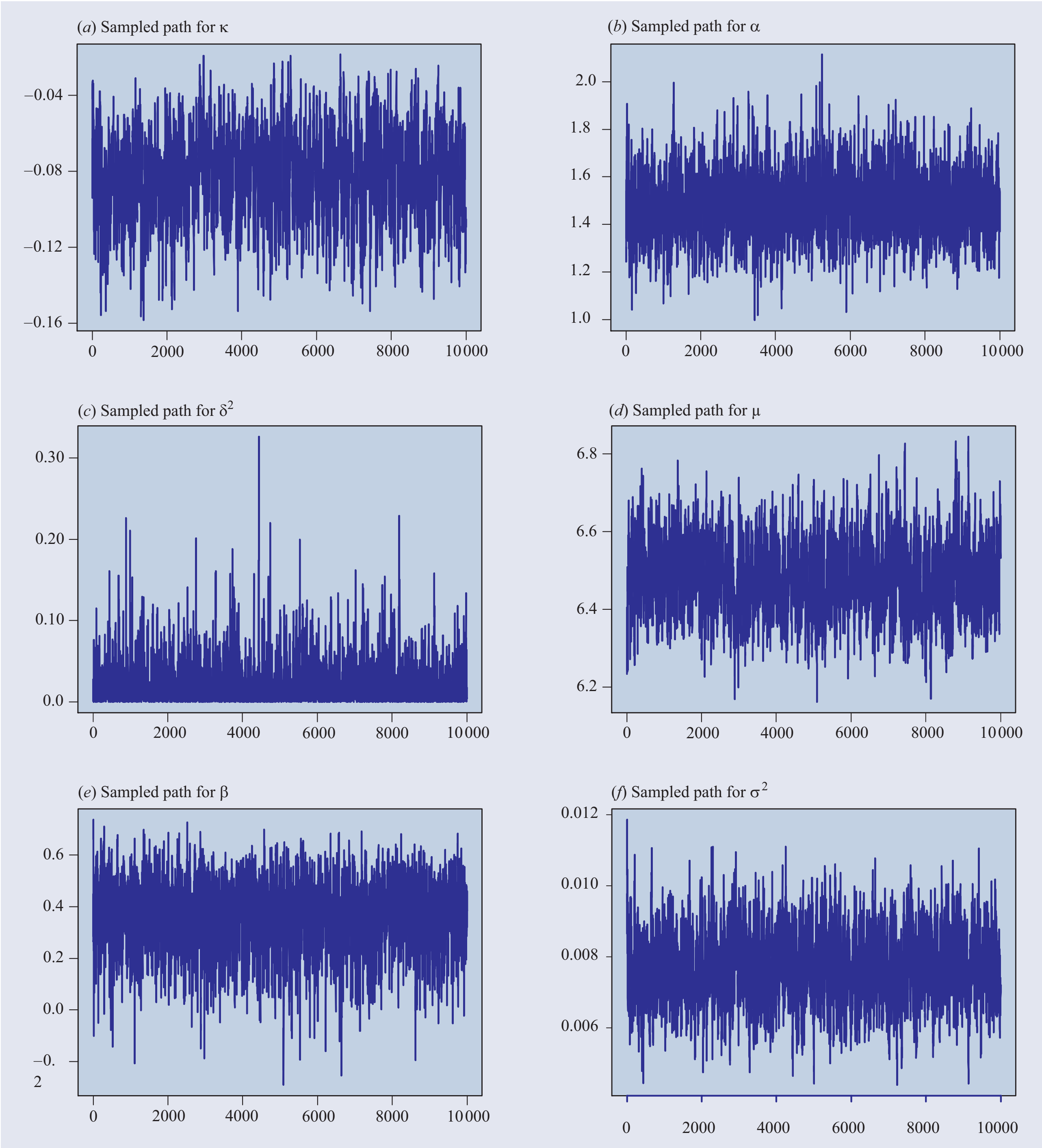

Figure 1. MCMC results for the Milstein scheme.

In addition to the batch-mean standard error, one may also compute the standard deviation $\tilde{\sigma}_f$ directly based on the sampled paths using the formula

$$\tilde{\sigma}_f = \left\{ \frac{1}{N-1} \sum_{i=1}^{N} [f(\theta^{[i]}) - \bar{f}_N]^2 \right\}^{1/2}. \qquad (21)$$

Kim *et al* (1998) indicated that the mixing performance of the sampled paths can be measured using the simulation inefficiency factor (SIF), also called the integrated autocorrelation time by Sokal (1996), which is estimated as the variance of the sample mean divided by the variance of the sample mean from a hypothetical sampler that draws independent random observations from the posterior distribution. Meyer and Yu (2000) showed that SIF is given by

$$\mathrm{SIF} = \frac{\hat{\sigma}_f^2}{\tilde{\sigma}_f^2}. \qquad (22)$$

In the empirical applications, the burn-in period is taken as 10 000 iterations and the number of total recorded iterations

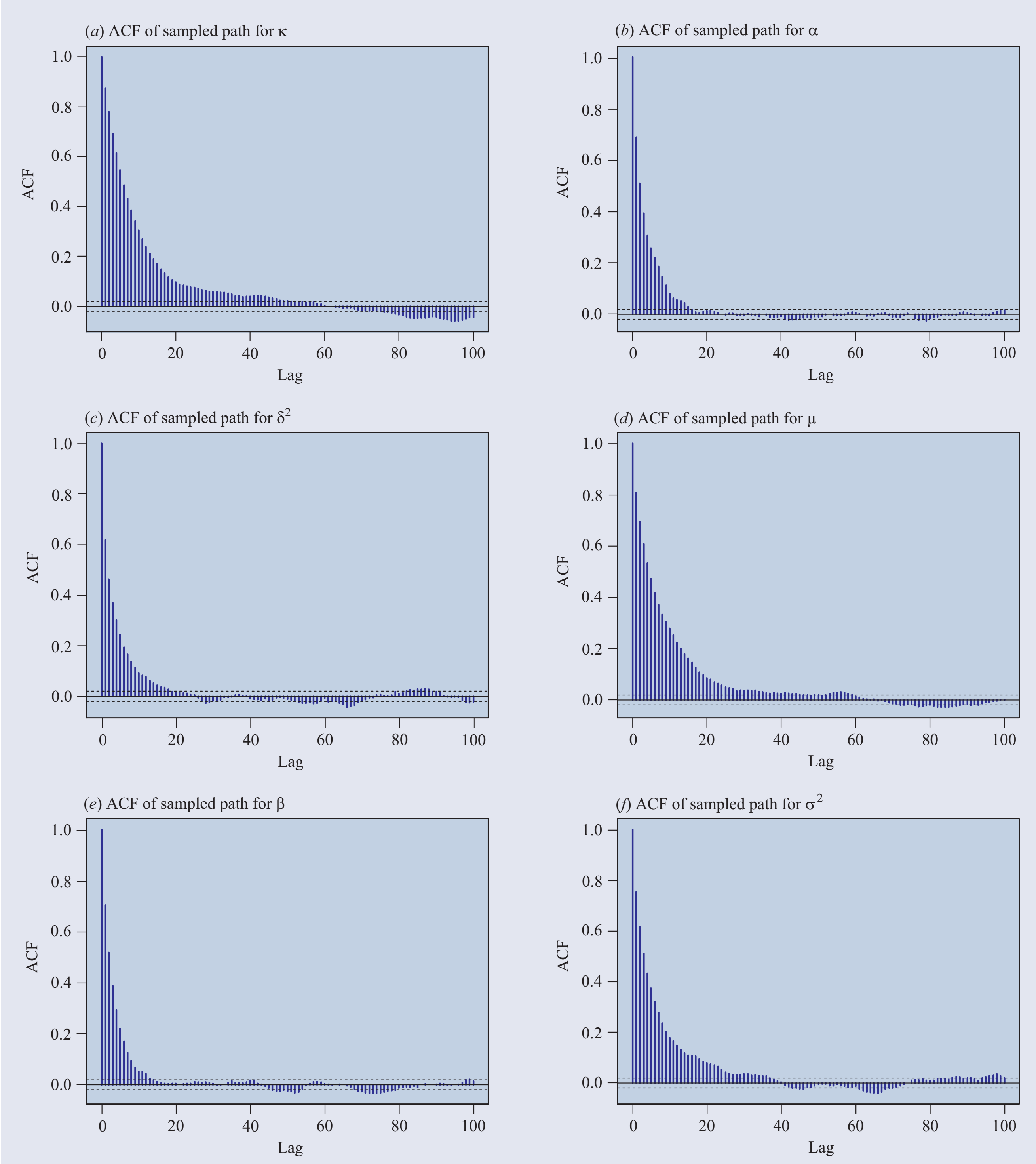

Figure 2. ACF for the Milstein scheme.

after the burn-in period is 50 000. Based on the sampled path for each data set, we calculate the ergodic average (or mean) and standard deviations. The MC standard errors are obtained using the batch-mean approach described in equations (17)–(19) with $f(x) = x$. The number of batches is $m = 50$, and there are $n = 1000$ draws in each batch.

We plot the MCMC sample paths of the parameters of the World index in figures 1(a)–(f), and the ACFs of these sample paths in figures 2(a)–(f)[7]. These plots show that the sample paths are reasonably well mixed. Table 1 summarizes the ergodic averages, standard deviations, 95% Bayes confidence intervals, Monte Carlo standard errors, and the SIFs for each data set. The Bayes confidence interval can be used to test the significance of each parameter. For example, for the

[7] To save space, the plots for the other two indexes are not presented.

Table 1. MCMC results of the Milstein scheme under specific priors. (Note: CI refers to the 95% confidence interval. SD refers to the standard deviation computed through (20). MCSE refers to the Monte Carlo standard error computed through the batch-mean approach. SIF refers to the SIF given by (22). AC refers to the acceptance probability.)

Data	Para.	Mean	CI	SD	MCSE	SIF	AC
MSCI World	κ	−0.081 30	(−0.127 32, −0.039 13)	0.022 25	0.000 91	82.85	0.20
	α	1.495 35	(1.256 44, 1.749 13)	0.123 35	0.003 56	41.70	0.24
	δ^2	0.020 58	(0.000 34, 0.093 82)	0.026 25	0.000 75	40.37	0.27
	μ	6.480 77	(6.308 44, 6.661 41)	0.090 40	0.003 58	78.44	0.21
	β	0.372 30	(0.081 79, 0.586 69)	0.127 43	0.003 14	30.28	0.22
	σ^2	0.007 65	(0.005 79, 0.009 58)	0.000 95	0.000 04	67.41	0.26
MSCI Europe	κ	−0.018 71	(−0.069 65, 0.028 70)	0.025 26	0.000 94	69.56	0.25
	α	1.563 59	(1.116 74, 1.815 34)	0.176 38	0.005 05	40.97	0.25
	δ^2	0.040 23	(0.000 57, 0.198 57)	0.053 89	0.002 39	98.58	0.24
	μ	6.316 99	(6.108 46, 6.548 61)	0.113 30	0.003 74	54.62	0.24
	β	0.272 17	(−0.214 48, 0.652 01)	0.221 47	0.006 74	46.37	0.25
	σ^2	0.010 43	(0.007 04, 0.013 47)	0.001 59	0.000 06	71.74	0.23
NYSE	κ	−0.035 71	(−0.080 39, 0.006 43)	0.022 18	0.000 79	63.95	0.23
	α	1.657 71	(1.298 26, 2.838 69)	0.136 35	0.003 08	25.46	0.23
	δ^2	0.014 00	(0.000 18, 0.068 29)	0.020 08	0.000 54	36.65	0.27
	μ	5.733 35	(5.547 33, 5.929 01)	0.096 67	0.003 59	68.83	0.24
	β	0.276 16	(−0.036 91, 0.512 57)	0.139 33	0.003 26	27.38	0.22
	σ^2	0.008 05	(0.006 16, 0.010 16)	0.001 02	0.000 04	64.42	0.24

World index all parameter estimates are significantly different from zero. Note that the posterior means of the steepness parameter α are quite similar across the three indexes. While the Europe index and the NYSE index are symmetrical (the sampled posterior β is not significantly different from zero), the World index is asymmetric. For the scale and volatility parameters (i.e. δ and σ), the World and NYSE indexes (but not the Europe index) have similar posterior means.

4.2.3. Robustness to the choice of the priors. To examine the robustness of the results with respect to the choice of the priors, we alter the priors in two ways. First, we keep the prior distributions in the same family as before but change some hyperparameters. The results are very similar. Second, we use a different set of prior distributions, which are now the uniform density. As any constant in the posterior will be eliminated from both the nominator and denominator when computing the acceptance probability, we effectively assume that the support of each uniform prior is wide enough for any update of the associated parameter. Then we use the same MCMC procedure as before and summarize the results in table 2.

A comparison with the results in table 1 reveals the following conclusions. Firstly, SIFs are either comparable (in most cases) to or marginally higher (in a few cases) than those under the priors adopted in section 4.2.1, suggesting that the mixing performance is not affected much by the change of priors. Secondly, there is no obvious difference in the ergodic averages and CIs under both sets of priors, suggesting that the posterior distribution is robust to the change of priors.

4.2.4. Empirical results under the Euler likelihood. To compare the MCMC performance based on the Milstein and Euler likelihoods, we apply the sampling algorithm to the posterior

$$\pi(\theta|\mathbf{x}) \propto \pi(\theta) p_E(\theta|\mathbf{x}), \tag{23}$$

where $\pi(\theta)$ is the same as that in section 4.2.3, and $p_E(\theta|\mathbf{x})$ is the Euler likelihood defined in (7). Applying the sampling algorithm to all three data sets using the same priors as adopted in section 4.2.1, we obtain the empirical results tabulated in table 3. The sample paths of parameters and the ACFs of sample paths are plotted, respectively, in figures 3(a)–(f) and figures 4(a)–(f).

A few more results emerge from table 3 and figures 3 and 4. First, the mixing performance under the Euler scheme is worse than that under the Milstein scheme, as can be seen from the fact that the sampled paths under the Euler scheme have larger variances than those under the Milstein scheme. This relative performance is also obvious from the plots. We apply the Heidelberger and Welch convergence test (Heidelberger and Welch 1983) to all the sampled paths, and find that the samples from the Milstein scheme under both sets of priors pass the test for all parameters, whereas the samples from the Euler scheme fail the test for δ^2. Second, the ergodic averages for some parameters are different from those obtained under the Milstein likelihood. For example, the estimated κ and β are, respectively, significantly different from the corresponding estimates reported in table 1. Better mixing and convergence under the Milstein scheme leads to the conclusion that the empirical results based on the Milstein likelihood are more reliable.

5. Empirical properties of hyperbolic diffusions

Rydberg (1999) reported simulation results of the normal inverse Gaussian diffusion in which the ACF of r_t^2 declines very slowly, thus partly satisfying dynamic property 2 listed in section 2. In this section we examine in more detail whether the hyperbolic diffusion would give rise to the statistical properties described in section 2.

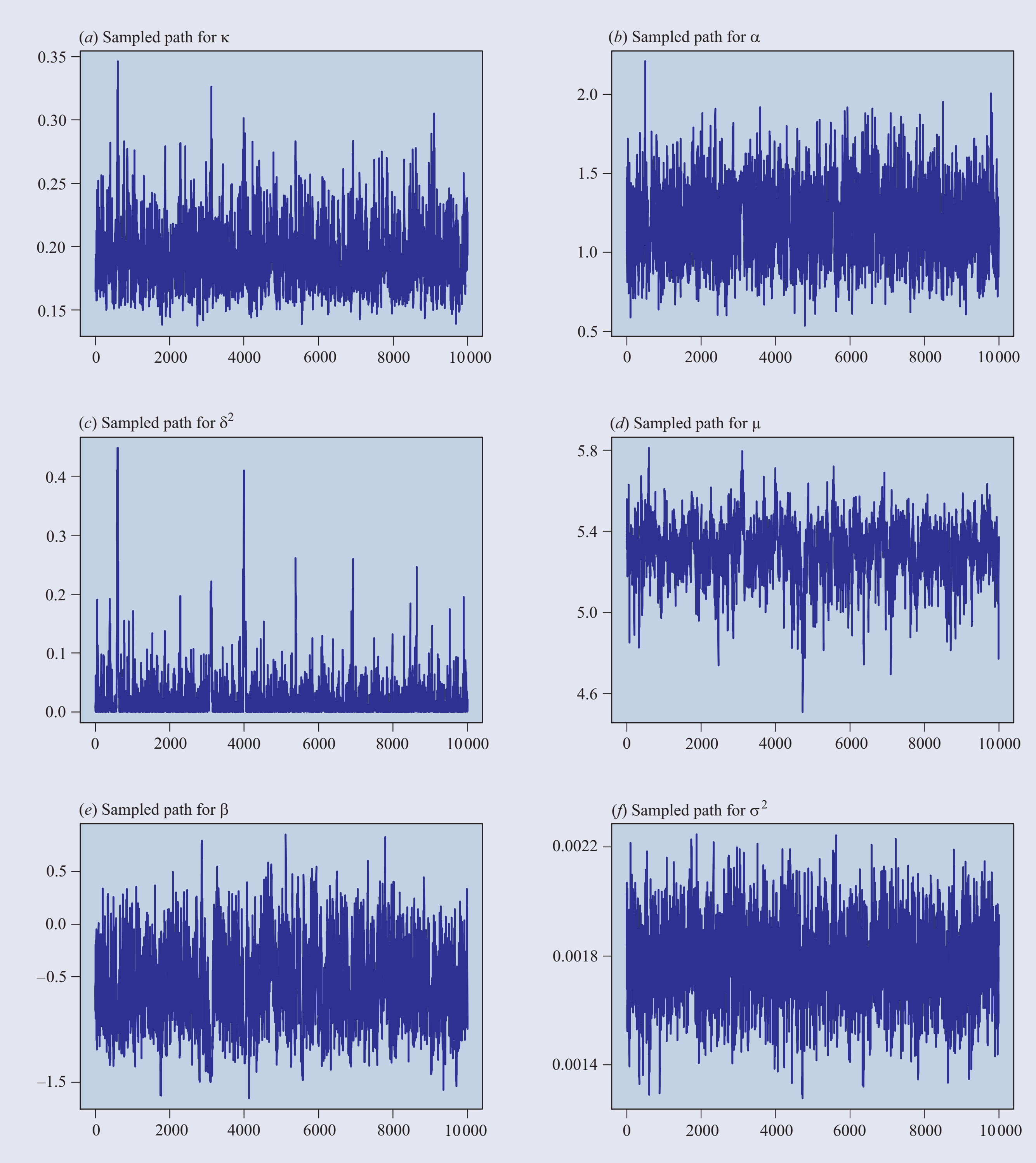

Figure 3. MCMC results for the Euler scheme.

To examine the posterior properties of the hyperbolic diffusion, we record 1000 sampled parameter vectors (i.e. 1 draw for every 50 draws of the MCMC iterations). Using the sampled parameter vector at each recorded draw, we generate a path of daily-price series with 2000 observations based on the Milstein approximation, where a time interval of 15 min is used (i.e. we use $\Delta t = 1/7000$ year, assuming 7 h of trading per day and 250 trading days per year). For each sampled path, we calculate the sample kurtosis, sample skewness, Box–Pierce statistic of the squared returns with 20 lags included, and ACF up to lag order 300 for $|r_t|^\phi$ with $\phi = 1, 1.5, 2$. Table 4 reports the means of the kurtosis, skewness, Box–Pierce statistic and the ACF of these 1000 sample paths. The last row of the table reports the proportions

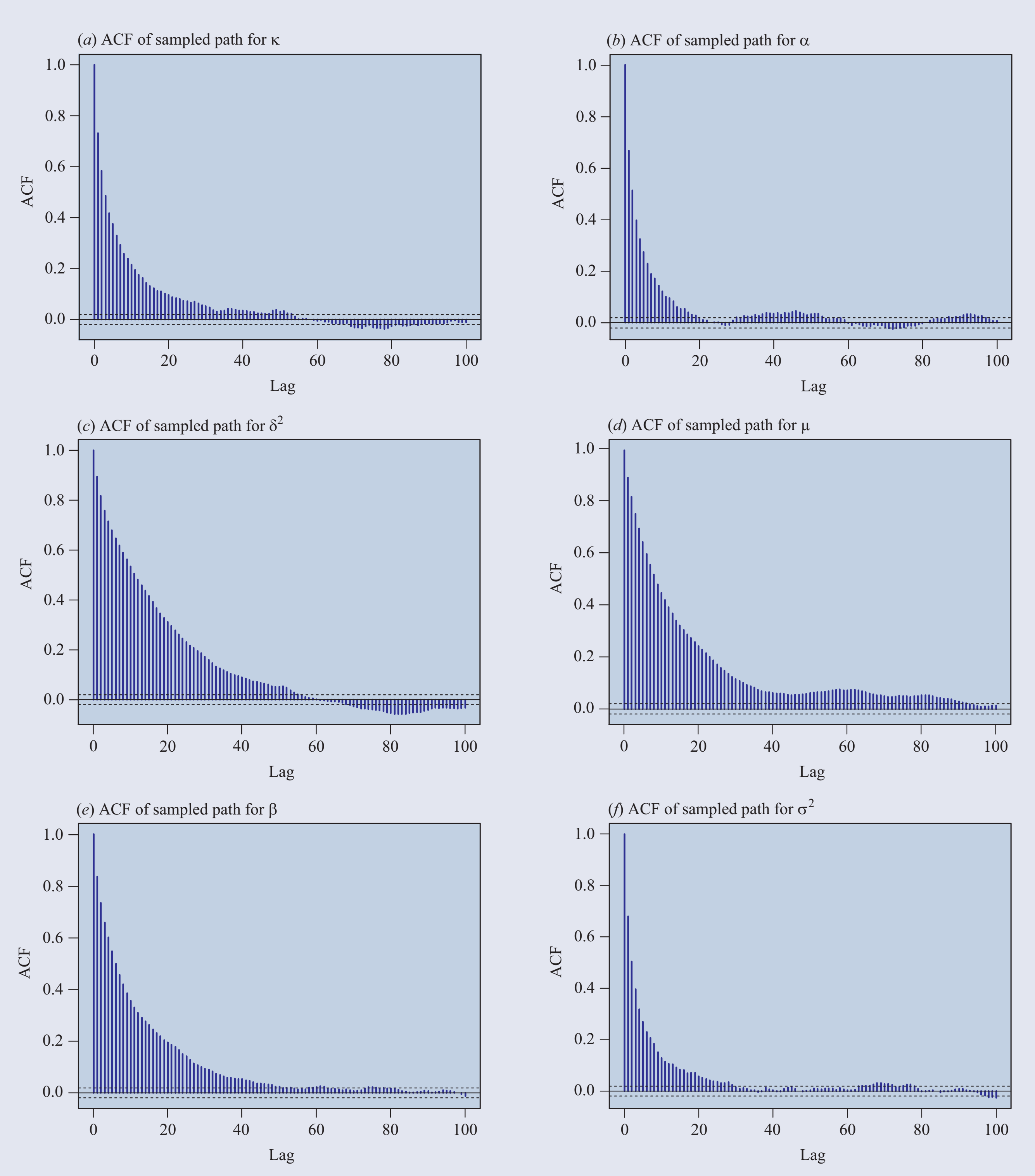

Figure 4. ACF for the Euler scheme.

that the kurtosis is larger than 3 (the kurtosis implied by the normal distribution), the skewness is less than 0, the Box–Pierce statistic is larger than 31.41 (the critical value at the 5% level), and the ACF at $\phi = 1$ is the largest among $\phi = 1, 1.5, 2$ at lags of order 100, 200 and 300 among the 1000 sample paths.

Several points can be observed from the table. First, there is overwhelming evidence of excess kurtosis (thick tails) in the unconditional distribution of the hyperbolic diffusion, as manifest in the mean of the kurtosis and the proportion of excess kurtosis across the simulated sample paths. This is consistent with static property 1. Second, there is weak evidence of asymmetry in the unconditional distribution: 44.3% of the time we get negative skewness whereas 55.7% of the time we have positive skewness. This is

Table 2. MCMC results of the Milstein scheme under uniform priors (note: see table 1).

Data	Para.	Mean	CI	SD	MCSE	SIF	AC
MSCI World	κ	−0.079 50	(−0.125 54, −0.035 52)	0.022 63	0.000 83	67.41	0.29
	α	1.539 33	(1.261 36, 1.859 25)	0.147 64	0.004 42	44.30	0.26
	δ^2	0.026 36	(0.000 37, 0.128 75)	0.036 35	0.001 53	87.95	0.27
	μ	6.482 24	(6.303 77, 6.663 00)	0.091 25	0.002 92	52.16	0.28
	β	0.357 68	(0.062 10, 0.588 73)	0.131 77	0.003 82	41.24	0.25
	σ^2	0.007 41	(0.005 13, 0.009 53)	0.001 09	0.000 04	62.65	0.26
MSCI Europe	κ	−0.014 22	(−0.068 31, 0.034 07)	0.025 53	0.001 00	76.17	0.27
	α	1.546 34	(0.995 61, 2.123 76)	0.272 17	0.012 47	102.84	0.29
	δ^2	0.051 39	(0.000 68, 0.242 00)	0.065 02	0.003 50	146.10	0.29
	μ	6.299 25	(6.067 92, 6.541 24)	0.122 31	0.003 91	51.06	0.27
	β	0.278 21	(−0.236 27, 0.738 44)	0.246 08	0.010 44	89.79	0.27
	σ^2	0.010 06	(0.005 84, 0.013 71)	0.001 99	0.000 10	134.66	0.26
NYSE	κ	−0.031 17	(−0.077 09, 0.013 61)	0.023 32	0.000 94	81.94	0.25
	α	1.657 71	(1.339 21, 2.048 63)	0.178 35	0.007 44	86.90	0.25
	δ^2	0.019 99	(0.000 24, 0.112 07)	0.030 67	0.001 52	119.81	0.26
	μ	5.730 78	(5.542 49, 5.923 04)	0.098 45	0.003 70	72.31	0.25
	β	0.231 27	(−0.115 84, 0.489 54)	0.154 36	0.005 11	54.95	0.27
	σ^2	0.007 66	(0.005 22, 0.010 03)	0.001 18	0.000 05	87.56	0.24

Table 3. MCMC results of the euler scheme under specific priors (note: see table 1).

Data	Para.	Mean	CI	SD	MCSE	SIF	AC
MSCI World	κ	0.187 43	(0.154 99, 0.240 44)	0.022 14	0.000 73	54.28	0.21
	α	1.168 44	(0.801 89, 1.603 88)	0.205 22	0.006 69	53.16	0.27
	δ^2	0.020 92	(0.000 20, 0.127 77)	0.037 68	0.001 73	105.95	0.22
	μ	5.303 67	(5.017 87, 5.539 62)	0.133 11	0.007 65	165.21	0.24
	β	−0.607 43	(−1.279 71, 0.198 68)	0.388 50	0.018 94	118.83	0.32
	σ^2	0.001 77	(0.001 50, 0.002 04)	0.000 13	0.000 01	45.80	0.22
MSCI Europe	κ	0.226 99	(0.193 02, 0.277 36)	0.021 13	0.001 13	141.94	0.21
	α	1.133 59	(0.739 79, 1.606 13)	0.222 33	0.010 26	106.52	0.27
	δ^2	0.016 59	(0.000 15, 0.105 10)	0.031 86	0.002 07	210.23	0.26
	μ	5.007 02	(4.392 87, 5.481 37)	0.255 09	0.018 98	276.91	0.22
	β	−0.032 55	(−1.085 18, 0.759 78)	0.444 67	0.032 83	272.52	0.24
	σ^2	0.001 59	(0.001 34, 0.001 85)	0.000 13	0.000 01	92.74	0.22
NYSE	κ	0.259 75	(0.210 57, 0.320 13)	0.028 03	0.000 86	46.62	0.29
	α	1.289 02	(0.910 02, 1.723 75)	0.208 30	0.009 35	100.73	0.29
	δ^2	0.051 59	(0.000 84, 0.218 60)	0.063 03	0.003 02	114.97	0.20
	μ	4.589 43	(4.339 53, 4.768 34)	0.111 07	0.004 99	100.91	0.28
	β	−1.115 54	(−1.599 22, −0.531 60)	0.263 77	0.012 20	107.02	0.29
	σ^2	0.002 83	(0.002 27, 0.003 30)	0.000 27	0.000 01	77.00	0.23

Table 4. Analysis of statistical properties of the hyperbolic diffusion.

				ACF of $\{\lvert r_t\rvert/\lvert r_t\rvert^{1.5}/\lvert r_t\rvert^2\}$		
	Kurt.	Skew	B–P	Lag 100	Lag 200	Lag 300
Mean	20.52	0.098	72.96	0.134/0.107/0.081	0.091/0.072/0.055	0.062/0.050/0.039
Prop.	1.00	0.443	0.668	0.736	0.732	0.624

consistent with static property 2. Third, the average value of the Box–Pierce statistic is much larger than the 5% critical value and the proportion of this statistic being significant is 67%, indicating reasonably strong evidence of an ARCH effect. Moreover, like Rydberg (1999), we also find that the ACF of $|r_t|$ and r_t^2 decay very slowly, a pattern inconsistent with the exponential decay. All these results are consistent with the dynamic property 2. Finally, at all three lags considered, most of the time the ACF is highest at power one. This result is consistent with the Taylor effect.

6. Conclusions

In this paper we propose a Bayesian MCMC method to estimate the hyperbolic diffusion based on the discretized density via the Milstein scheme. Relative to some alternative estimation methods, such as the ML estimation based on the discretized densities and the MCMC method based on the discretized density via the Euler scheme, we find that the MCMC method using the Milstein scheme provides the best empirical results. Apart from showing evidence that the MCMC method is a useful tool for estimating hyperbolic diffusions and making statistical inferences, we have also demonstrated that the hyperbolic diffusion is able to exhibit many of the stylized facts about asset returns documented in the financial econometrics literature.

Acknowledgment

XZ would like to acknowledge the research support from the National Science Foundation of China.

JY gratefully acknowledges financial support from the Royal Society of New Zealand Marsden Fund under number 01-UOA-015. Some of the work was done when JY was visiting the National University of Singapore in 2001, whose hospitality is gratefully acknowledged.

References

Barndorff-Nielsen O E 1978 Hyperbolic distributions and distributions on hyperbolae *Scand. J. Statistics* **5** 151–7
Bibby B M and Sorensen M 1995 Martingale estimation functions for discretely observed diffusion processes *Bernoulli* **1** 17–39
Bibby B M and Sorensen M 1997 A hyperbolic diffusion model for stock prices *Finance Stochastics* **1** 25–41
Carr P and Wu L 2003 Time-changed Levy processes and option pricing *J. Financial Economics*, at press
Chib S 2001 Markov chain Monte Carlo methods: computation and inference *Handbook of Econometrics* vol 5, ed J J Heckman and E Leamer (Amsterdam: North-Holland) pp 3569–649
Chibs S and Greenberg E 1995 Understanding the Metropolis–Hastings algorithm *Am. Statistician* **49** 327–35
Cox J, Ingersoll J and Ross S 1985 A theory of the term structure of interest rates *Econometrica* **53** 385–407
Elerian O 1998 A note on the existence of a closed-form conditional transition density for the Milstein scheme *Economics Discussion Paper* 1998-W18 Nuffield College, Oxford
Elerian O, Chib S and Shephard N 2001 Likelihood inference for discretely observed non-linear diffusions *Econometrica* **69** 959–93
Engle R F 1982 Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation *Econometrica* **50** 987–1007
Eraker B 2001 MCMC analysis of diffusion models with application to finance *J. Business Economic Statistics* **19** 177–91
Eraker B, Johannes M and Polson N 2003 The impact of jumps in volatility and returns *J. Finance* **58** 1269–300
Geweke J 1999 Using simulation methods for Bayesian econometric models: inference, development, and communication *Econometric Rev.* **18** 1–73
Gilks W R, Richardson S and Spiegelhatler D J 1996 Introducing Markov chain Monte Carlo *Markov Chain Monte Carlo in Practice* ed W R Gilks, S Richardson and D J Spiegelhalter (London: Chapman and Hall) pp 45–57
Heidelberger P and Welch P 1983 Simulation run length control in the presence of an initial transient *Operations Res.* **31** 1109–44
Heston S L 1993 A closed-form solution for options with stochastic volatility, with application to bond and currency options *Rev. Financial Studies* **6** 327–43
Johannes M and Polson N 2003 MCMC method for financial econometrics *Handbook of Financial Econometrics* ed Y Aït-Sahalia and L Hansen at press
Kim S, Shephard N and Chib S 1998 Stochastic volatility: likelihood inference and comparison with ARCH models *Rev. Econ. Stud.* **65** 361–93
Kloedan P E and Platten E 1995 *Numerical Solution of Stochastic Differential Equations* (New York: Springer)
Kou S G 2002 A jump-diffusion model for option pricing *Management Sci.* **48** 1086–101
Meyer R and Yu J 2000 BUGS for a Bayesian analysis of stochastic volatility models *Econometrics J.* **3** 198–215
Milstein G N 1978 A method of second-order accuracy integration of stochastic differential equations *Theory Probab. Appl.* **23** 396–401
Robert C P and Casella G 1999 *Monte Carlo Statistical Methods* (New York: Springer)
Roberts G O 1996 Markov chain concepts related to sampling algorithms *Markov Chain Monte Carlo in Practice* ed W R Gilks, S Richardson and D J Spiegelhalter (London: Chapman and Hall) pp 45–57
Rydberg T H 1999 Generalized hyperbolic diffusion processes with applications in finance *Math. Finance* **9** 183–201
Ryden T, Teräsverta T and Åsbrink S 1998 Stylized facts of daily return series and the hidden Markov model *J. Appl. Econometrics* **13** 217–44
Sokal A D 1996 Monte Carlo methods in statistical mechanics: foundations and new algorithms *Lectures at the Cargese summer school on 'Functional Integration: Basics and Applications'*