Estimation of Continuous Time Processes Via the Empirical Characteristic Function

George J. Jiang* and John L. Knight†

This Version: August 2000

*George J. Jiang, Department of Finance, Schulich School of Business, York University, 4700 Keele Street, Toronto, Ontario, Canada M3J 1P3. Tel: (416) 736-2100 ext. 33302, (416) 736-5073 and Fax: (416) 736-5687. E-mail: gjiang@schulich.yorku.ca.

†John L. Knight, Department of Economics, University of Western Ontario, London, Ontario, Canada, email: jknight@julian.uwo.ca.

Estimation of Continuous Time Processes Via the Empirical Characteristic Function

Abstract

This paper examines a particular class of continuous-time stochastic processes commonly known as affine diffusions (AD) and affine jump-diffusions (AJD) in which the drift, the diffusion and the jump coefficients are all affine functions of the state variables. By deriving the joint characteristic function associated with a vector of observed state variables for such models, we are able to examine the statistical properties of these diffusions and jump-diffusions as well as develop an efficient estimation technique based on empirical characteristic functions (ECF) and a GMM estimation procedure based on *exact* moment conditions. The estimators developed in this paper are in stark contrast to those available in the literature in the sense that our methods require neither discretization nor simulation. We demonstrate that our methods are in particular useful for the AD and AJD models with latent variables, i.e. the case where some of the state variables are unobserved. We illustrate our approach with a detailed examination of the continuous-time square-root stochastic volatility (SV) model, along with an empirical application using S&P 500 index returns.

JEL Classification: C13, C22, C52, G10

Key Words: Affine Diffusion, Affine Jump-Diffusion, Empirical Characteristic Function (ECF), Generalized Method of Moments (GMM), Stochastic Volatility (SV)

1 Introduction

The estimation of continuous-time stochastic processes, especially those with unobserved or latent variables, has posed a challenge to statisticians and econometricians for some time. In financial applications, due to the unavailability of a continuous sample of observations, estimation has usually been performed by first discretizing the model and then applying various moment based estimation methods, e.g. the generalized method of moment (GMM) approach in Chan, Karolyi, Longstaff, and Sanders (1992). In the econometrics and statistical literature, new estimation techniques are mostly developed based on simulation methods, e.g. the simulated method of moments (SMM) in Duffie and Singleton (1993), and the efficient method of moments (EMM) in Gallant and Tauchen (1996). The application of these approaches has had varying success due mainly to the need of both discretizing the model and simulating sample paths. Beside the intensive computation involved, these two steps can compound the estimation errors and consequently may lead to poor finite sample properties. While there are a few specific models for which the maximum likelihood (ML) estimation is possible as there exist explicit closed form transition density functions, these models have not proved to be popular in finance due to unrealistically simplistic model specifications. In the multivariate framework, the estimation problem becomes even more difficult and particularly so if some of the state variables are unobserved.

However, it has recently be noticed that there is a class of continuous-time stochastic processes where, while the transition density functions are not known, their Fourier transformation, i.e. the characteristic functions are known. These processes are the so-called affine diffusions (AD) and affine jump-diffusions (AJD) developed and popularized in a series of papers by Duffie and Kan (1996), Dai and Singleton (1999), and Duffie, Pan and Singleton (1999). Since these processes are Markovian with, in many cases, explicit closed form conditional characteristic functions, it opens the door for alternative estimation techniques. Two recent papers which have exploited the idea of developing new estimation methods based on conditional characteristic functions are Chacko and Viceira (1999) and Singleton (1999).

This paper also examines the AD and AJD models via their associated characteristic functions. However, in our case we use the unconditional joint characteristic function rather than the conditional characteristic function. Based on the unconditional joint characteristic function, we examine the statistical properties of these models and develop new estimation strategies. In particular, our interest centers on models where some of the state variables are unobserved. The estimation of these models, which include the popular stochastic volatility model as a special case, poses even more of a challenge to statisticians and econometricians. Singleton (1999) proposes the use of the conditional characteristic function along with simulation and develops a SMM estimator. Our approach exploits the explicit functional form of the unconditional joint characteristic function to develop an efficient estimation procedure based on the empirical joint characteristic function. Since there is an exact one-to-one correspondence between the characteristic function and the distribution function, the empirical characteristic function (ECF) contains the same amount of information as the empirical distribution function (EDF). Consequently, the ECF method has the same asymptotic efficiency as the maximum likelihood (ML) method. Moreover, since the *exact* moment conditions of the stochastic processes are readily available from the analytical characteristic functions, an alternative approach is also proposed in this paper for the AD and AJD models via the generalized method of moments (GMM). The estimators developed in this paper are in stark contrast to those available in the literature in the sense that our methods require neither discretization nor simulation. More importantly, our methods are particularly useful for the AD and AJD models with latent variables, i.e. the case where some of the state variables are unobserved. We illustrate the approach of analyzing the statistical properties of AD and AJD models based on the joint characteristic function with two examples, the square-root process in the univariate case and the SV model in the bivariate case. An application of the estimation methods is undertaken for the SV model using the S&P 500 index returns

Section 2 examines the general AD and AJD models and the joint characteristic function in both the situation where all the state variables are observed and that where only some of the state variables are observed. We show how from the joint characteristic function we can examine the statistical properties of these processes. In section 3 we consider the estimation problem and develop, in particular, two estimators, the empirical characteristic function (ECF) approach and the GMM approach for the AD and AJD models with unobserved or latent variables. Section 4 involves an empirical application of our techniques to the SV model using daily S&P 500 index returns, along with further discussion on the issues related to the implementation of the ECF method. A brief conclusion is contained in Section 5.

2 The Continuous Time Stochastic Processes

We consider a general continuous time affine jump-diffusion (AJD) model, with affine diffusion (AD) model as a special case, as defined in Duffie, Pan and Singleton (1999). Using the same notation as in Duffie, Pan and Singleton (1999), we fix a probability space $(\Omega, \mathcal{F}, P)$ and an information filtration $(\mathcal{F}_t) = \{\mathcal{F}_t : t \geq 0\}$, and suppose that X_t is a Markov process in some state space $\mathcal{D} \in \mathbb{R}^n$, following the stochastic differential equation (SDE):

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t + dZ_t \tag{1}$$

where W_t is an $(\mathcal{F}_t)$-standard Brownian motion in $\mathbb{R}^n$, $\mu(\cdot) : \mathcal{D} \to \mathbb{R}^n$ and $\sigma(\cdot) : \mathcal{D} \to \mathbb{R}^n$ are respectively the drift function and diffusion function, and Z is a pure jump process whose jumps have a fixed probability distribution $\mathcal{J}$ on $\mathbb{R}^n$ and arrive with intensity $\{\lambda(X_t) : t \geq 0\}$, for some $\lambda(\cdot) : \mathcal{D} \to [0, \infty)$. The initial value of the stochastic process X_0 is assumed to follow a trivial distribution. For X_t to be a well-defined Markov process, regularity conditions on the filtration $(\mathcal{F}_t) = \{\mathcal{F}_t : t \geq 0\}$ and restrictions on the state space as well as on the coefficient functions of the stochastic process, namely $(\mathcal{D}, \mu(\cdot), \sigma(\cdot), \lambda(\cdot), \mathcal{J})$, are required.

For technical details, see e.g. Ethier and Kurtz (1986), Duffie, Pan and Singleton (1999), Dai and Singleton (1999), and Duffie and Kan (1996).

The continuous time jump-diffusion (JD) process defined in (1) is often used in the finance literature to model the dynamics of asset return, index return, exchange rate, and interest rate, etc. Intuitively, the drift term $\mu(\cdot)$ represents an instantaneous deterministic time trend of the process, the diffusion term $\sigma(\cdot)\sigma(\cdot)'$ represents an instantaneous volatility of the process when no jump occurs, and the jump term Z_t captures the discontinuous change of the sampling path with both random arrival of jumps and random jump sizes. More specifically, as noted in Duffie, Pan and Singleton (1999), conditional on the path of X_t, the jump times of Z_t are the jump times of a Poisson process with time varying intensity $\{\lambda(X_s) : 0 \leq s < t\}$, and the size of the jump of Z_t at a jump time τ is independent of $\{X_s : 0 \leq s < \tau\}$ and follows the probability distribution $\mathcal{J}$.

For convenience and tractability, many financial models impose an "affine" structure on the coefficient functions $\mu(\cdot), \sigma(\cdot)\sigma(\cdot)'$, and $\lambda(\cdot)$, i.e. all of these functions are assumed to be affine on $\mathcal{D}$. Using the notation in Duffie, Pan and Singleton (1999), we have

$$\mu(X_t) = K_0 + K_1 X_t,$$

$$[\sigma(X_t)\sigma(X_t)']_{ij} = [H_0]_{ij} + [H_1]_{ij} X_t,$$

$$\lambda(X_t) = l_0 + l_1 X_t$$

where $K = (K_0, K_1) \in \mathbb{R}^n \times \mathbb{R}^{n\times n}, H = (H_0, H_1) \in \mathbb{R}^{n\times n} \times R^{n\times n\times n}, l = (l_0, l_1) \in \mathbb{R} \times \mathbb{R}^n$. Let $g(c) = \int_{\mathbb{R}^n} \exp\{c \cdot z\} d\mathcal{J}(z)$ be the jump transform whenever the integral is well defined, where $c \in \mathbb{C}^n$ the set of n-tuples of complex numbers, $g(\cdot)$ determines the jump size distribution. It is obvious that the set of "coefficients" or parameters (K, H, l, g) completely specifies the AJD process and determines its statistical properties, given the initial condition X_0. When the jump intensity is set as zero, i.e. $\lambda(\cdot) = 0$, the process is referred to as the affine diffusion (AD) process.

Dynamic properties of the jump-diffusion (JD) process defined in (1) are determined by the transition density functions which, under regularity conditions, satisfy both the Kolmogorov forward and backward (or Fokker-Planck) equations of the Markov process. However, the transition density functions of the diffusion and jump-diffusion process in general do not have a closed analytical form. For instance, in the simplest univariate pure-diffusion case, i.e. $n = 1, \lambda(\cdot) = 0$, the Ornstein-Ulenbeck process and the square-root diffusion process are the only two well-known processes with explicit transition density functions. For these two processes, both the drift function $\mu(\cdot)$ and the diffusion function $\sigma^2(\cdot)$ are of linear structure. The functional forms of the transition densities corresponding to specifications essentially different from the above two processes are not known explicitly, see Wong (1964).

As we mentioned in the introduction, since there is an exact one-to-one correspondence between the characteristic function and the distribution function, an alternative expression of the statistical properties of the model is given by the characteristic function of the state variables. Duffie, Pan and Singleton (1999) showed that under the "affine" structure, the conditional characteristic function of the jump-diffusion process as defined in (1) has a closed form. Using their results, we have the characteristic function of $X_{t+\tau}$ conditional on X_t given by

$$\begin{aligned} \psi(u; X_{t+\tau}|X_t) &= E[\exp\{iuX_{t+\tau}\}|X_t] \\ &= \exp\{C(\tau,u) + D(\tau,u)'X_t\} \end{aligned} \tag{2}$$

where $D(\cdot)$ and $C(\cdot)$ satisfy the following complex-valued Ricatti equations:

$$\frac{\partial D(\tau,u)}{\partial \tau} = K_1' D(\tau,u) + \frac{1}{2} D(\tau,u)' H_1 D(\tau,u) + l_1(g(D(\tau,u)) - 1) \tag{3}$$

$$\frac{\partial C(\tau,u)}{\partial \tau} = K_0' D(\tau,u) + \frac{1}{2} D(\tau,u)' H_0 D(\tau,u) + l_0(g(D(\tau,u)) - 1) \tag{4}$$

with boundary conditions: $D(0,u) = iu, C(0,u) = 0$. With certain specification of the coefficient function (K, H, l, g), explicit solutions of $D(\cdot)$ and $C(\cdot)$ can be found. In other cases, as noted in Duffie, Pan and Singleton (1999), the solution would have to be found numerically. With explicit solutions of $D(\cdot)$ and $C(\cdot)$, it appears that it would be easier to investigate the statistical properties of the model in the frequency domain than in the time domain.

In practice, in order to investigate the joint sampling properties of the jump-diffusion process on the one hand and eliminate the "start-up" effect of the initial condition on the other, the unconditional joint probability distribution or the unconditional joint characteristic function of the state variables $\{X_{t_1}, X_{t_2}, ..., X_{t_p}\}$ over time $\{t_1, t_2, ..., t_p\}$ rather than the conditional probability density function or conditional characteristic function are often required. In the next section, we will see that these joint distribution functions or joint characteristic functions will be needed for the estimation of the stochastic processes. Due to the Markov property, the joint probability distribution is given by the conditional probability distribution and the marginal distribution of the initial condition, namely $f(X_{t_1}, X_{t_2}, ..., X_{t_p}) = f(X_{t_1}) \prod_{i=1}^{p-1} f(X_{t_{i+1}}|X_{t_i})$. Obviously the joint density function cannot have closed form if the transition density function is not explicitly known. Again, our hope is to resort to the unconditional joint characteristic function. Even though the characteristic function does not share the nice iterative property as the distribution function, the following lemma shows that in an affine jump-diffusion framework, when the marginal characteristic function of the initial condition is available, we still can have a closed form for the joint characteristic function.

Lemma 1: Let $y = (x_1, x_2, ..., x_{p+1})$ be a vector of $p + 1$ equi-spaced observations on the affine jump-diffusion process X_t as defined in (1) with initial condition $X_0 = x_0$. The joint characteristic function of this vector conditional on x_0 is given by

$$\begin{aligned}\psi(u_1, u_2, ..., u_{p+1}; x_1, x_2, ..., x_{p+1}|x_0) &= E[\exp\{i \sum_{k=1}^{p+1} u_k' x_k\}|x_0] \\ &= \exp\{\sum_{k=1}^{p+1} C(1, u_k^*)\} \exp\{D(1, u_1^*)' x_0\}\end{aligned}$$

where $u_{p+1}^* = u_{p+1}$ and $u_k^* = u_k - iD(1, u_{k+1}^*), k = 1, ..., p$. If the characteristic function of

x_0 is available, i.e.

$$E[e^{u'x_0}] = \phi(u; x_0)$$

then, the unconditional characteristic function of y is given by

$$\psi(u_1, u_2, ..., u_{p+1}; x_1, x_2, ..., x_{p+1}) = \exp\{\sum_{k=1}^{p+1} C(1, u_k^*)\}\phi(D(1, u_1^*); x_0) \tag{5}$$

Proof: See Appendix.

Remark 1: In the case that X_t is stationary, the marginal density function and the marginal characteristic function of the process can be derived from the conditional density function and the conditional characteristic function respectively, namely $f(X_t = x_t) = \lim_{\tau \to +\infty} f(X_t = x_t | X_{t-\tau} = x_{t-\tau})$ and $\phi(u; X_t) = \lim_{\tau \to +\infty} \psi(u; X_t | X_{t-\tau} = x_{t-\tau})$. Statistical properties, both static and dynamic, can be derived from the characteristic functions as both the cumulants and moments can be calculated easily when the characteristic function has a closed analytical form.

As an application of the above lemma and an illustration of how the characteristic function can be used to study the statistical properties of the stochastic process, we now examine the following example.

Example 1: The Square-Root Diffusion Process. Consider the following univariate pure diffusion process

$$dV_t = \beta(\alpha - V_t)dt + \sigma V_t^{1/2} dW_t \tag{6}$$

with affine structure on both the drift function $\mu(V_t) = \beta(\alpha - V_t)$ and diffusion function $\sigma^2(V_t) = \sigma^2 V_t$. This process is known to be a positive process with a reflecting barrier at zero, which is attainable when $2\beta\alpha < \sigma^2$, and has been used in Cox, Ingersoll and Ross (1985) to model the nominal interest rates and Bailey and Stulz (1989) and Heston (1993) to model the conditional volatility of asset returns. Solving from (3) and (4) for $D(\cdot)$ and $C(\cdot)$, we have the conditional characteristic function

$$\begin{aligned}\psi(u; V_t = v_t | V_0 = v_0) &= \exp(C(1,u) + D(1,u)v_0) \\ &= \exp\{-\frac{2\beta\alpha}{\sigma^2}\ln(1-\frac{iu}{c}) + \frac{iue^{-\beta t}}{1-iu/c}v_0\} \end{aligned} \quad (7)$$

where $c = 2\beta/(\sigma^2(1-e^{-\beta t}))$. Via Fourier inversion of the conditional characteristic function, the transition density function can be obtained as $f(V_t = v_t | V_0 = v_0) = ce^{-u-\nu}(\frac{\nu}{u})^{q/2} I_q(2(u\nu)^{1/2})$ with V_t taking nonnegative values, where $u = cv_0e^{-\beta t}$, $\nu = cv_t$, $q = \frac{2\beta\alpha}{\sigma^2} - 1$, and $I_q(\cdot)$ is the modified Bessel function of the first kind of order q. The transition density function is non-central chi-square, $\chi^2[2cv_t; 2q+2, 2u]$, with $2q+2$ degrees of freedom and parameter of noncentrality $2u$ proportional to the current level of the stochastic process. The conditional expected value and variance of V_t is given by $E[V_t|V_0 = v_0] = v_0e^{-\beta t} + \alpha(1-e^{-\beta t}), Var[V_t|V_0 = v_0] = \frac{v_0\sigma^2}{\beta}(e^{-\beta t} - e^{-2\beta t}) + \frac{\alpha\sigma^2}{2\beta}(1-e^{-\beta t})^2$. If the process displays the property of mean reversion ($\beta > 0$), the process is stationary and its marginal distribution can be derived from the transition density, which is a gamma probability density function, i.e., $f(V_t = v_t) = \frac{\omega^s}{\Gamma(s)}v_t^{s-1}e^{-\omega v_t}$ where $\omega = 2\beta/\sigma^2$ and $s = 2\alpha\beta/\sigma^2$, with mean α and variance $\alpha\sigma^2/2\beta$. Similarly the marginal characteristic function can be derived from the conditional characteristic function as $\phi(u; V_t) = (1 - \frac{iu\sigma^2}{2\beta})^{-2\beta\alpha/\sigma^2}$.

With the direct application of Lemma 1, the joint characteristic function of $\{v_1, v_2, ..., v_{p+1}\}$ as a vector observations of V_t with initial condition $V_0 = v_0$ is given by

$$\begin{aligned}\psi(u_1, u_2, ..., u_{p+1}; v_1, v_2, ..., v_{p+1}) &= E[E[\exp\{i\sum_{k=1}^{p+1} u_k v_k\}]|v_0] \\ &= \exp\{\sum_{k=1}^{p+1} C(1, u_k^*)\}E[\exp\{D(1, u_1^*)v_0\}]\end{aligned}$$

where $u_{p+1}^* = u_{p+}, u_k^* = u_k - iD(1, u_{k+1}^*)$ for $k = 1, 2, ..., p$ with $D(\cdot)$ and $C(\cdot)$ given by the solutions in (7) for this particular example. Now since v_0 has a gamma marginal (stationary) distribution, we have

$$\psi(u_1, u_2, ..., u_{p+1}; v_1, v_2, ..., v_{p+1}) = \exp\{\sum_{k=1}^{p+1} C(1, u_k^*)\}(1 - \frac{D(1, u_1^*)\sigma^2}{2\beta})^{-2\beta\alpha/\sigma^2}$$

A widely used and also very important class of stochastic processes in finance literature is the one that involves the unobservable or latent state variables, e.g. the stochastic volatility (SV) model. As we will see in the next section, the estimation of such a class of models has posed a great challenge to econometricians. We partition the whole vector of X_t into two sub-vectors, i.e. $X_t = (S_t', V_t')'$, where $S_t \in \mathbb{R}^m, n > m > 0$, is the vector of observed state variables and $V_t \in \mathbb{R}^{n-m}$ is the vector of unobserved or latent state variables. Instead of finding the joint sampling properties of the whole vector of state variables X_t, now we are interested in only the sub-vector of X_t, namely S_t. While we know that the vector process X_t is Markovian, S_t, as a sub-vector of X_t, is not necessarily Markovian. That is, the future values of S_t may be dependent upon not only its current observation, but also its historical observations, and in certain cases even its entire history. In the non-Markov case, the joint probability function $f(s_{t_0}, s_{t_1}, ..., s_{t_p})$ does not have the nice iterative property any more. Again, when its analytical functional form is not available, we hope to resort to the joint characteristic function. Specializing earlier results, we have

$$\begin{aligned} &\psi(u1, u2; s_{t+\tau}, v_{t+\tau}|s_t, v_t) = E[\exp\{iu1's_{t+\tau} + iu2'v_{t+\tau}\}|s_t, v_t] \\ = \ &\exp\{C(\tau, u1, u2) + D1(\tau, u1, u2)'s_t + D2(\tau, u1, u2)'v_t\} \end{aligned} \tag{8}$$

where $C(\tau, u1, u2) = C(\tau, u), (D1(\tau, u1, u2)', D2(\tau, u1, u2)') = D(\tau, u)'$ with $u = (u1', u2')'$. Further, we note

$$\begin{aligned} \psi(u; s_{t+\tau}|s_t, v_t) &= E[\exp\{iu's_{t+\tau}\}|s_t, v_t] \\ &= \exp\{C(\tau, u, 0) + D1(\tau, u, 0)'s_t + D2(\tau, u, 0)'v_t\} \end{aligned}$$

Therefore we have the following lemma.

Lemma 2: Let X_t be the affine jump-diffusion process defined in (1) with $X_t = (S_t', V_t')'$, and $y = (s_1, s_2, ..., s_{p+1})$ be a vector of $p+1$ equi-spaced observations on the partial stochastic process S_t with initial condition $X_0 = x_0$. The joint characteristic function of the vector

y conditional on x_0 is given by

$$\psi(u1_1, u1_2, ..., u1_{p+1}; s_1, s_2, ..., s_{p+1}|x_0) = E[\exp\{i\sum_{k=1}^{p+1} u1'_k s_k\}|x_0]$$
$$= \exp\{\sum_{k=1}^{p+1} C(1, u1^*_k, u2^*_k)\} \exp\{D(1, u1^*_1, u2^*_1)' x_0\}$$

where $u1^*_{p+1} = u1_{p+1}, u2^*_{p+1} = 0$, $u1^*_k = u1_k - iD1(1, u1^*_{k+1}, u2^*_{k+1})$, and $u2^*_k = -iD2(1, u1^*_{k+1}, u2^*_{k+1})$ for $k = 1, 2, ..., p$. If the characteristic function of x_0 is available, i.e.

$$E[e^{u' x_0}] = \phi(u; x_0)$$

Then, the unconditional joint characteristic function of y is given by

$$\psi(u1_1, u1_2, ..., u1_{p+1}; s_1, s_2, ..., s_{p+1})$$
$$= \exp\{\sum_{k=1}^{p+1} C(1, u1^*_k, u2^*_k)\}\phi(D(1, u1^*_1, u2^*_1); x_0) \quad (9)$$

Proof: See Appendix.

Remark 2: As noted in Remark 1, for stationary processes the marginal density function can be derived from the transition density, and the marginal characteristic function can be derived from the conditional characteristic function. In the asset return models, it is quite often the case that not the whole vector of X_t is stationary. Instead, only part of the process is stationary, and the rest of the process is first difference stationary. For instance, in the following stochastic volatility asset return model, while the stochastic volatility is specified to be stationary, the logarithmic asset prices are modeled as first difference stationary, in other words, the asset returns are stationary. Without loss of generality, we assume that V_t is stationary and $\Delta S_t = S_t - S_{t-\Delta}$ is stationary, that is $(\Delta S_t, V_t)$ is a stationary process. Instead of analyzing the joint characteristic function of $\{s_1, s_2, ..., s_{p+1}\}$, we can derive the joint characteristic function of $\{\Delta s_1, \Delta s_2, ..., \Delta s_{p+1}\}$, in which only the marginal characteristic function of v_0 is needed. For S_t to be first difference stationary, certain restrictions on the parameter set (K, H, l, g) are required. One set of sufficient conditions is that the elements of

K_1, H_1, l_1 and g corresponding to the S_t equations are all zero, i.e. the state variables S_t do not appear directly in its own dynamic process. Thus the solution to the SDE (3) for $D(\tau, u)$, corresponding to state variables S_t, is $D(\tau, u) = D(0, u) = iu$. We illustrate Lemma 2 in general and this special case in particular with the following example.

Example 2: The Square-Root Stochastic Volatility Model. Consider the following asset return process with stochastic conditional volatility for the logarithmic asset prices s_t [1]

$$\begin{aligned} ds_t &= \mu dt + V_t^{1/2} dW_t \\ dV_t &= \beta(\alpha - V_t)dt + \sigma V_t^{1/2} dW_t^v \\ dW_t dW_t^v &= \rho dt, \quad t \in [0, T] \end{aligned} \tag{10}$$

This continuous-time SV model has been widely used in the finance literature for asset return dynamics as it allows for a closed-form solution for European option prices. As in Singleton (1999), the drift term of the asset return process is specified as a constant. It is noted that when the drift term of the asset return process is specified as a linear function of the state variable V_t, both the European option prices and the conditional characteristic function of the asset return will still yield closed forms. The specification of the instantaneous volatility process in the above model guarantees the nonnegativeness of the volatility. The solution of the square-root process in (6) or (10) can be written as

$$V_{t+1} = (1 - e^{-\beta})\alpha + e^{-\beta} V_t + \int_t^{t+1} \sigma e^{-\beta(t+1-\tau)} V_\tau^{1/2} dW_\tau^v$$

which is of an $AR(1)$ form, where $\epsilon_{t+1} = \int_t^{t+1} \sigma e^{-\beta(t+1-\tau)} V_\tau^{1/2} dW_\tau^v$ is a martingale. Thus the process can be viewed as an $AR(1)$ process with heteroskedasticity. The parameter β measures the intertemporal persistence of the volatility process, and the correlation between dW_t^v and dW_t measures the level of asymmetry of the conditional volatility. In particular, when $\rho < 0$ we have the so-called "leverage effect".

[1] Please note that in this example, s_t denotes the random process.

Statistical properties of the stochastic volatility process have been discussed in detail in Example 1. Due to the lack of explicit solutions to the continuous-time asset return process with stochastic volatility as defined in (10), there has been no formal derivation or discussion of the statistical properties of the asset return process. However, from a straightforward application of Lemma 2 to the process $(\Delta s_t, V_t)$, we can derive the unconditional characteristic function of the asset return $\Delta s_t = s_t - s_{t-1}$, as well as the joint characteristic function of the asset returns. Lemma 3 presents the joint characteristic function of asset returns, followed by the discussion of both static and dynamic properties.

Lemma 3: Let s_t be the stochastic process defined in (10), given s_0 and that $V_0 = v_0$ follows its marginal distribution, the joint characteristic function of $\Delta s_1 = s_1 - s_0, \Delta s_2 = s_2 - s_1, ...,$ and $\Delta s_{p+1} = s_{p+1} - s_p$, where $p \geq 1$, i.e.,

$$\psi(u_1, u_2, ..., u_{p+1}; \Delta s_1, \Delta s_2, ..., \Delta s_{p+1}) = E[\exp\{i \sum_{k=1}^{p+1} u_k \Delta s_k\}]$$

can be derived as

$$\begin{aligned} & \ln \psi(u_1, u_2, ..., u_{p+1}; \Delta s_1, \Delta s_2, ..., \Delta s_{p+1}) \\ = & \sum_{k=1}^{p+1} C(1; u1_k^*, u2_k^*) - \frac{2\beta\alpha}{\sigma^2} \ln(1 - \frac{\sigma^2}{2\beta} \sum_{k=1}^{p+1} D(1; u1_k^*, u2_k^*)) \end{aligned} \quad (11)$$

where $u1_{p+1}^* = u1_{p+1}, u2_{p+1}^* = 0, u1_k^* = u_k, u2_k^* = -iD(1; u1_{k+1}^*, u2_{k+1}^*)$ for $k = 1, ..., p$ and

$$\begin{aligned} C(\tau; u_1, u_2) &= (iu_1\mu + i\beta\alpha u_2)\tau + \frac{\beta\alpha}{\sigma^2}[(b-h)\tau - 2\ln(\frac{1 - ge^{-h\tau}}{1-g})] \\ D(\tau; u_1, u_2) &= \frac{b-h}{\sigma^2} \cdot \frac{1 - e^{-h\tau}}{1 - ge^{-h\tau}} \end{aligned}$$

with $h(u_1, u_2) = [b^2 + \sigma^2(u_1^2 + 2\rho\sigma u_1 u_2 + \sigma^2 u_2^2 + 2\beta u_2 i)]^{1/2}, g(u_1, u_2) = (b-h)/(b+h), b = \beta - \rho\sigma u_1 i - \sigma^2 u_2 i$. In particular, when $p = 1$, we have

$$\begin{aligned} \ln \psi(u_1, u_2; \Delta s_1, \Delta s_2) &= C(1; u_2, 0) + C(1; u_1, -iD(1; u_2, 0)) \\ & -\frac{2\beta\alpha}{\sigma^2} \ln(1 - \frac{\sigma^2(D(1; u_2, 0) + D(1; u_1, -iD(1; u_2, 0)))}{2\beta}) \end{aligned}$$

Proof: See Appendix.

The static properties of the model can be derived from the marginal characteristic function, which is in essence a special case of the joint characteristic function in Lemma 3 with $p = 0$. That is, we have

$$\begin{aligned} \phi(u; \Delta s_t) &= E[\exp\{iu\Delta s_t\}] \\ &= \exp\{C(1, u, 0)\}(1 - \frac{\sigma^2 D(1, u, 0)}{2\beta})^{-2\alpha\beta/\sigma^2} \end{aligned} \tag{12}$$

From (12) and the expressions of $C(1, u, 0)$ and $D(1, u, 0)$ we see immediately that if $\rho = 0$ then the characteristic function $\phi(u; \Delta s_t)$ is given by

$$\phi(u; \Delta s_t) = e^{iu\mu} Q(u^2; \alpha, \beta, \sigma^2)$$

where $Q(u^2; \alpha, \beta, \sigma^2)$ is a function of u^2 and hence is real. Thus Δs_t has a symmetric distribution around its mean μ.

Remark 3: Further from the marginal characteristic function of Δs_t in (12), we note that the asset return process defined in (10) is a stationary process with the following first four moments

$$\begin{aligned} E[\Delta s_t] &= \mu, \\ Var[\Delta s_t] &= \alpha, \\ E[(\Delta s_t - \mu)^3] &= \frac{3}{\beta^2}(e^{-\beta} + \beta - 1)\alpha\rho\sigma, \\ E[(\Delta s_t - \mu)^4] &= 3Var[\Delta s_t]^2 + \phi_0 \end{aligned} \tag{13}$$

where $\phi_0 = \frac{3}{\beta^3}(e^{-\beta} + \beta - 1 + 4((2+\beta)e^{-\beta} + \beta - 2)\rho^2)\alpha\sigma^2$. Note that the third moment can be non-zero due to the presence of the asymmetric parameter ρ. It is obvious that the asset return distribution is asymmetric if $\rho \neq 0$, and the skewness has the same sign as that of ρ. From the fourth moment, we can see that $E[(\Delta s_t - \mu)^4] - 3Var[\Delta s_t]^2 = \phi_0 > 0$, i.e. the asset return distribution has positive excess kurtosis, or fat tails. It can be further shown that the distribution of asset returns Δs_t is leptokurtic, more peaked in the vicinity of its mean than the distribution of a comparable normal random variable. These features are consistent with the empirical findings on the unconditional distributions of many financial asset returns.

Remark 4: From the joint characteristic function in Lemma 3, we can readily calculate various cross moments of Δs_t. In particular, we have (i) the mean adjusted asset return $\Delta s_t - \mu$ is uncorrelated over time, i.e.

$$E[(\Delta s_t - \mu)(\Delta s_{t+\tau} - \mu)] = 0, \quad \tau > 1$$

and (ii) the squared mean adjusted asset return $(\Delta s_t - \mu)^2$ is correlated over time with,

$$\begin{aligned} & Cov[(\Delta s_t - \mu)^2, (\Delta s_{t+\tau} - \mu)^2] \\ = \; & \frac{1}{2\beta^3} e^{-(\tau+1)\beta}(e^\beta - 1)(e^\beta - 1 + 4\rho^2(e^\beta - \beta - 1))\alpha\sigma^2, \quad \tau > 1 \end{aligned} \tag{14}$$

which is strictly positive but decreases with τ and goes to 0 as τ goes to $+\infty$. Thus, while the asset returns are uncorrelated over time, the squared asset returns are autocorrelated. We also note that the squared return process can behave quite differently than the instantaneous volatility process.

The discussion and derivations surrounding Example 2 above highlights the fact that often in affine diffusions with latent variables can we exploit the form of the characteristic functions, both conditional and unconditional, to examine the statistical properties of these processes. In the next section, we will see that the joint characteristic function can also be used to develop efficient estimators of the parameters in AD and AJD models, especially for the case that some of the state variables are unobserved or latent.

3 GMM and Empirical Characteristic Function (ECF) Estimation of the Continuous Time Stochastic Processes

In this section, we focus on the estimation of the affine jump-diffusion process as defined in (1) that involves latent variables, especially the case that the observed partial process is non-Markovian. The processes without latent variables can be viewed as special cases, which have been intensively studied in Singleton (1999). For the case that the conditional characteristic function of the affine diffusion or affine jump-diffusion process X_t has explicit analytical form and the whole vector of X_t is observable, Singleton (1999) proposes both the maximum likelihood (ML) estimation and partial maximum likelihood (PML) estimation based on Fourier inversion of the conditional characteristic function (CCF), as well as the standard quasi-maximum likelihood (QML) estimation based on conditional moments. Singleton (1999) also proposes both the efficient and approximately efficient estimation of the AD or AJD process X_t based directly on the empirical conditional characteristic function (ECCF). As will be seen later in this section, the basic idea is to match the analytical CCF to the empirical CCF. For the efficient estimation, however, the optimal weight function is not known as it cannot be computed without *a prior* knowledge of the conditional density function, which we will also discuss in detail later in this section. When part of the state variables of X_t is unobserved or latent, Singleton (1999) proposes the simulated method of moments (SMM) estimation based on the CCF, which involves simulating the stochastic process. Since the model is essentially in continuous time, such simulation often involves approximation error due to the discretization of the data generating process (DGP). Chacko and Viceira (1999) construct a GMM estimator based on the unconditional mean of the difference between the empirical characteristic function and analytical characteristic function for integer values of the dummy variable u. [2] However, their procedure does not utilize

[2] We wish to thank the anonymous referees for bringing our attention to this research.

the full conditioning information in the data, and therefore while their estimation example is truly exact, the lack of conditioning information produces an estimator that is neither as efficient as that in Singleton (1999) nor as the estimators we shall propose in this section.

Estimation of dynamic nonlinear latent variable models, such as the SV model in (10), is by no means a trivial task. The difficulty arises due to the fact that the latent variables, i.e. the stochastic volatility in the case of the SV model, are unobservable, and thus the models cannot be estimated using standard maximum likelihood (ML) methods as the latent or unobserved variable has to be integrated out of the likelihood.[3] Over the past few years, however, remarkable progress has been made in the field of statistics and econometrics regarding the estimation of nonlinear latent variable models in general and SV models in particular. Various estimation methods have been proposed for SV models, which are mostly simulation-based, computationally intensive and involve discretization when applied to the continuous-time processes. For example, we have the Quasi Maximum Likelihood (QML) by Harvey, Ruiz and Shephard (1994), the Monte Carlo Maximum Likelihood by Sandmann and Koopman (1997), the Markov Chain Monte Carlo (MCMC) methods by Jacquier, Polson and Rossi (1994) and Kim, Shephard and Chib (1998) for discrete-time SV models, and the Efficient Method of Moments (EMM) by Gallant and Tauchen (1996) for both discrete-time and continuous-time SV models. Applications of EMM have been performed by Gallant, Hsieh and Tauchen (1997), Andersen and Lund (1997) to symmetric continuous-time SV models, and Andersen, Benzoni and Lund (1998) and Chernov and Ghysels (1999) to asymmetric continuous-time SV models. To our knowledge, there have been very few attempts to estimate the continuous-time model as specified in (10) with non-zero correlation between asset returns and conditional volatility (i.e. $\rho \neq 0$). In addition to the aforementioned

[3]This is not a standard problem since the dimension of this integral equals the number of observations, which is typically large in financial time series. Standard Kalman filter techniques cannot be applied due to the fact that either the latent process is non-Gaussian or the resulting state-space form does not have a conjugate filter.

EMM applications, Singleton (1999) applies the proposed simulation based ECF estimator for affine diffusion models to the SV model as defined in (10) using the S&P 500 index returns. Chacko and Viceira (1999) apply their GMM estimation to the jump-diffusion process with stochastic volatility using both weekly and monthly total return observations on the CRSP value-weighted portfolio.

In this paper, we exploit the fact that the characteristic function of the asset return process as well as the joint asset return process can often be derived analytically. Using these analytical results, we propose the following two different methods for the estimation of the continuous-time SV process, one is the generalized method of moments (GMM) based on the *exact* moment conditions and the other is the empirical characteristic function (ECF) method which matches the empirical characteristic function calculated from the data to the analytical characteristic function derived from the model. As in Singleton (1999), for the estimators discussed in this section, we assume that Hansen's (1982) regularity conditions are satisfied.

3.1 GMM Estimation of the Continuous Time Stochastic Processes

As shown in Section 2, various unconditional moments of the state variables can be derived from the unconditional joint characteristic function. These moments are *exact* in the sense that they are corresponding to the continuous time DGP without any discretization or approximation. When the joint characteristic function has a closed analytical form, these moment conditions also have explicit expressions whenever they exist. Consequently, these moment conditions can be used to estimate the stochastic processes following a standard GMM procedure. In this section, we present this procedure for the case that the partial process S_t of X_t is observed and first difference stationary, and the rest of the process is latent. For other cases, the only difference is the change of notation. Let $\psi(u_1, .., u_t, .., u_{t+j}, ..u_{p+1}; \Delta s_1, .., \Delta s_t, .., \Delta s_{t+j}, .., \Delta s_{p+1})$ be the unconditional joint characteristic function of $\{\Delta s_1, .., \Delta s_t, ..,$

$\Delta s_{t+j}, .., \Delta s_{p+1}\}$ with $1 \leq t \leq t+j \leq p+1$, we have

$$\begin{aligned} & E[\Delta s_t^{k_1} \Delta s_{t+j}^{k_2}] \\ = \; & \frac{\partial^{k_1+k_2} \psi(u_1, .., u_t, .., u_{t+j}, .., u_{p+1}; \Delta s_1, .., \Delta s_t, .., \Delta s_{t+j}, .., \Delta s_{p+1})}{i^{k_1+k_2} \partial^{k_1} \Delta s_t \partial^{k_2} \Delta s_{t+j}} |_{u_1 = ... = u_{p+1} = 0} \end{aligned}$$

where $k_1, k_2 \in \{0, 1, ...\}$. Further, we define

$$f_t(\theta) = \begin{bmatrix} \Delta s_t^{k_1^a} \Delta s_{t+j}^{k_2^a} - E[\Delta s_t^{k_1^a} \Delta s_{t+j}^{k_2^a}] \\ \Delta s_t^{k_1^b} \Delta s_{t+j}^{k_2^b} - E[\Delta s_t^{k_1^b} \Delta s_{t+j}^{k_2^b}] \\ ... \end{bmatrix}$$

$$k_1^a, k_1^b, k_2^a, k_2^b \in \{0, 1, ...\}; t, t+j \in \{1, ..., p+1\}.$$

as the vector of various chosen moment conditions.

With the vector of the population moment conditions denoted above by $f_t(\theta)$, and the vector of corresponding sample moment conditions with sample size T denoted by $g_T(\theta)$, the GMM estimator is defined as

$$\hat{\theta}_T = \arg\min_{\theta \in \Theta} \{\mathcal{J}_T(\theta) = g_T'(\theta) \mathcal{W}_T(\theta) g_T(\theta)\} \tag{15}$$

where Θ denotes the permissible parameter space and $\mathcal{W}_T$ a positive definite weight matrix which is chosen to yield the smallest asymptotic covariance matrix of the GMM estimator of θ as in Hansen (1982). Under regularity conditions, the estimator $\hat{\theta}_T$ is consistent and asymptotically normal, i.e.

$$T^{1/2}(\hat{\theta}_T - \theta) \sim AN(0, V_T) \tag{16}$$

and a consistent estimator of V_T is given by $\hat{V}_T = \frac{1}{T}(\hat{D}'(\theta)\hat{S}^{-1}(\theta)\hat{D}'(\theta))^{-1}$, where $\hat{D}(\theta)$ is the Jacobian matrix of $g_T(\theta)$ with respect to θ evaluated at the estimated parameters, and $\hat{S}(\theta)$ is a consistent estimator of $S(\theta) = E[f_t(\theta) f_t'(\theta)]$. Under the null hypothesis that the model is true, the minimized value of $\mathcal{J}_T(\theta)$ in (15) is χ^2 distributed with degrees of freedom equal to the number of orthogonality conditions net of the number of parameters to be estimated. This χ^2 statistic provides a goodness-of-fit test for the model, and a high value of this statistic suggests that the model is misspecified.

3.2 ECF Estimation of the Continuous Time Stochastic Processes

Before we discuss the estimation of the AD and AJD model via the ECF, it is worthwhile to briefly outline the ECF estimation method for stationary stochastic processes. Suppose a vector of random variables y has the distribution function $F(y;\theta)$ where $\theta \in \Theta$ the parameter space which specifies the distribution, the characteristic function is defined as the Fourier transformation of the distribution function. In particular, the joint characteristic function of y, denoted by $\psi(u;y,\theta)$, is defined as

$$\psi(u;y,\theta) = E[\exp\{iu'y\}] = \int \exp\{iu'y\}dF(y;\theta) \tag{17}$$

and the joint empirical characteristic function, denoted by $\psi_n(u;y)$, of the random vector y is the sample counterpart of the characteristic function,

$$\psi_n(u;y) = \frac{1}{n}\sum_{j=1}^{n}\exp\{iu'y_j\} = \int \exp\{iu'y\}dF_n(y;\theta) \tag{18}$$

where $F_n(y)$ is the empirical distribution function. Since there is an unique Fourier-Stieltjes transformation, the joint characteristic function and the joint empirical characteristic function defined above contain the same amount of information as the distribution function and the empirical distribution function (i.e. the sampling observations) respectively. The basic idea of the empirical characteristic function estimation method for time series models was first proposed by Feuerverger (1990) by extending his earlier results on the ECF method (see Feuerverger and McDunnough (1981) and Feuerverger and Mureika (1977)) to estimate stationary stochastic processes. He proposed splitting the data into $n = T - p$ overlapping blocks, each of length $p + 1$, and using the joint characteristic function of each block to estimate the parameter vector θ by minimizing the following weighted distance [4]

$$\min_{\theta} \int ... \int \mid \psi(u;y,\theta) - \psi_n(u;y) \mid^2 \varphi(u)du \tag{19}$$

[4]Alternatively, one can define the distance as a summation of the squared difference between the analytical characteristic function and empirical characteristic function over a discrete set of u.

or, under standard regularity conditions, solving the first order conditions of the above optimization problem

$$\int ... \int (\psi(u; y, \theta) - \psi_n(u; y))w(u, \theta)du = 0 \tag{20}$$

where $\psi(u; y, \theta)$ and $\psi_n(u; y)$ are given in (17) and (18) and $y_j = (x_j, x_{j+1}, ..., x_{j+p})$, $j = 1, 2, ..., n = T - p$ with x_j being observations on the time series. Both $\varphi(\cdot)$ and $w(\cdot)$ are weight functions. Since (20) is the first order conditions of (19), for these two procedures to be equivalent it is necessary for $w(\cdot)$ to be a function of θ as it involves the derivative of $\psi(\cdot)$ with respect to θ. As $\psi(u; y, \theta)$ is a function of the unknown parameters, i.e., θ and $\psi_n(u; y)$ a function of the data, the ECF method merely minimizes (19) with respect to θ. Feuerverger (1990) also showed that using (20) with $w(u, \theta)$ chosen as

$$w(u, \theta) = \frac{1}{(2\pi)^{p+1}} \int \cdots \int \frac{\partial \ln f(x_j | x_{j-1}, ..., x_{j-p})}{\partial \theta} \exp(-iu_j' y_j) dy_j$$

results in an estimator asymptotically equivalent to maximum likelihood. Lemma 4 gives the asymptotic distribution associated with the ECF estimator derived from (19).

Lemma 4. The parameter estimators $\hat{\theta}$ obtained from the empirical characteristic function method is consistent and asymptotically normally distributed with

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Omega) \tag{21}$$

where $\Omega = B^{-1}(\theta)A(\theta)B^{-1}(\theta)$ is the variance-covariance of the parameter estimators with both $A(\theta)$ and $B(\theta)$ defined in the Appendix, and a consistent estimator of the variance-covariance is also given in the Appendix.

Proof: See Appendix.

In section 2 we noted that for the AD and AJD models, the joint characteristic function often has a closed form. This expression can be used in the above outlined estimation strategy to develop relatively efficient estimators of the unknown parameters.

In the full Markov model where all variables are observed, our proposed estimator will be an alternative to those proposed in Singleton (1999) and Chacko and Viceira (1999). Both these papers use the conditional characteristic function rather than the joint characteristic function, with the Singleton (1999) approach making full use of the conditioning information. The major difference between our approach and the approach in Singleton (1999) is that instead of conditioning on the initial observation, our approach includes the initial observation in the joint unconditional distribution. With large sample size and for the Markov models, the approaches would be equivalent as it is clear that we would simply need blocks of length two. Therefore, in situations where all variables in the model are observed we recommend the use of the Singleton (1999) GMM-CCF estimator or our estimator with block size of two, or $p = 1$. Singleton's (1999) estimator uses a finite grid of points for the estimation of the conditional characteristic function and his estimator hinges on the choice of the number of points and their value. Our estimator with $p = 1$ only requires the choice of the weight function $\varphi(u)$. How these estimators compare in finite samples and how they perform for different models are important and interesting issues. However, these issues are beyond the scope of this paper and will be examined in a separate Monte Carlo study.

As we mentioned in the introduction, the more challenging estimation problem arises for the AD and AJD models where some of the variables are unobserved or latent. In this case, the difficulty arises with the estimation methods based on the conditional characteristic function as not all the conditioning information is observed. These unobserved state variables therefore need to be integrated out. The issue is further complicated when the observed state variables, as part of the whole state variables which are Markovian, are not necessarily Markovian. That is, the future states may depend on not only the current states but also their past and in some cases even the entire history. This situation can be handled with our estimator and was also a subject of discussion in both Chacko and Viceira (1999) and Singleton (1999). Chacko and Viceira's (1999) strategy is to first examine the conditional characteristic function and then to integrate out the conditioning information resulting in the marginal or unconditional characteristic function, from which moment conditions and a GMM estimator are derived. Unfortunately, the marginal characteristic function is unable to help explain any time series behavior in the observed process. Singleton's (1999) approach relies on simulation methods to integrate out the dependence of the conditional characteristic function on the unobserved or latent variables. However, it could be argued that the use of simulation along with discretization of the continuous-time process seems to negate the tractability arguments associated with the affine AD and AJD models in terms of their characteristic functions.[5]

The proposed estimators in this paper, both the GMM and ECF, exploit the known form of the joint unconditional characteristic function. As shown in Section 2, the joint unconditional characteristic function often has closed analytical form for the AD and AJD models, thus (19) or (20) can be readily used for the estimation of the processes. One of the advantages of the estimators proposed in this paper is that they are *exact* approaches, in the sense that they require neither discretization nor simulation. More specifically, in the GMM case we are able to use both marginal and joint moment conditions resulting in a parsimonious, powerful and intuitive GMM procedure. The ECF estimator is also appealing as it essentially performs GMM with a continuum of moment conditions and has the added property of relative efficiency. When the weight function is asymptotically optimal, the ECF method has the same asymptotic efficiency as the ML estimation method. Furthermore, the joint unconditional characteristic function reflects both the static and time series behavior of the observed process. As will be discussed later in this section, for Markov process, it is sufficient to have block size of two, or $p = 1$, in the estimation. While for non-Markov process, it would appear that there is no general solution to the choice of the block size.

We will conclude this section by examining in some detail the Singleton (1999) empirical conditional characteristic function (ECCF) estimator for Markov models, showing how it

[5]We thank one of the referees for this argument.

can be extended to non-Markov models and how easily it can be adapted to generate the exact ML estimator rather than the conditional ML estimator. More specifically, Singleton's (1999) results exploit the particular dependency in Markov models resulting in the following optimal estimating equations, similar to (20), which result in conditional MLE:

$$\sum_{j=2}^{T} \int w(u,\theta|x_{j-1})(e^{iux_j} - \psi(u|x_{j-1}))du = 0$$

where

$$w(u,\theta|x_{j-1}) = \frac{1}{2\pi} \int \frac{\partial \ln f(x_j|x_{j-1})}{\partial \theta} e^{iux_j} dx_j$$

and

$$\psi(u|x_{j-1}) = E[e^{iux_j}|x_{j-1}].$$

The following Lemma extends some of the ideas in Singleton (1999) to non-Markov models and gives the optimal estimating equations which result in exact Maximum Likelihood estimators.

Lemma 5. For a general stationary stochastic process with the conditional likelihood given by

$$L \approx \prod_{j=2}^{n} f(x_j|I_{j-1})$$

where I_{j-1} signifies information up to and including time $j-1$, the optimal weight function associated with observation j, is given by

$$w_j(u,\theta|I_{j-1}) = \frac{1}{2\pi} \int e^{-iux_j} \frac{\partial \ln f(x_j|I_{j-1})}{\partial \theta} dx_j \tag{22}$$

and the resulting estimating equations leading to the conditional MLE are:

$$\sum_{j=2}^{n} \left(\int w_j(u,\theta|I_{j-1}) e^{iux_j} du \right) = 0 \tag{23}$$

For exact MLE we need to amend the estimating equations to take into account the additional information in the marginal pdf of x_1, i.e., $f(x_1)$. The estimating equations become:

$$\int w_1(u,\theta) e^{iux_1} dx_1 + \sum_{j=2}^{n} \int w_j(u,\theta|I_{j-1}) e^{iux_j} dx_j = 0 \tag{24}$$

with

$$w_1(u,\theta) = \frac{1}{2\pi}\int e^{-iux_1}\frac{\partial \ln f(x_1)}{\partial \theta}dx_1 \tag{25}$$

Proof: See Appendix.

If the process is Markov, then

$$f(x_j|I_{j-1}) = f(x_j|x_{j-1})$$

and unlike in the general case, this pdf maintains a constant form since the information set is now just the $(j-1)^{th}$ observation. Consequently, the weight function will also be of constant form, viz.,

$$w_j(u,\theta|I_{j-1}) = w_2(u,\theta|x_{j-1}), \quad \forall j > 2$$

When the conditional density is unknown but the process is Markov we can obtain an approximation to the optimal weight function by approximating the $\ln f(x_j|x_{j-1})$, e.g. via an Edgeworth expansion. This approach is currently under investigation by the authors.

However, when the process is not Markov and $f(x_j|I_{j-1})$ is not known we cannot use the Edgeworth approach as the form of $f(x|I_{j-1})$ is not constant. Consequently, we have no alternative but to use some simulation approach along with the ECF as in Singleton (1999) or use the weighted function of the integrated MSE given by (19). Using (19) raises the issues of the dimension of the joint characteristic function, i.e., the length of the overlapping blocks given by $(p+1)$ and the choice of the weight function $\varphi(u)$.

Unfortunately, it would appear that there is no general solution to the choice of p or $\varphi(u)$, except perhaps in a Markov model where $p = 1$ is clearly sufficient to capture the dependency in the data. Consequently, optimal values for p and $\varphi(u)$ need to be considered on a case by case basis.

4 An Empirical Application: Estimation of the Stochastic Volatility (SV) Model

In this section, an empirical application of the proposed methods in section 3, namely the ECF method and GMM, is undertaken for the SV model as defined in (10), namely

$$\begin{aligned} ds_t &= \mu dt + V_t^{1/2} dW_t \\ dV_t &= \beta(\alpha - V_t)dt + \sigma V_t^{1/2} dW_t^v \\ dW_t dW_t^v &= \rho dt, \quad t \in [0, T] \end{aligned} \tag{26}$$

The data set in our application consists of the daily S&P 500 index returns over the period from 1990 to 1999, obtained from DataStream International. The summary statistics of the static and dynamic properties of daily S&P 500 index returns are reported in Table 1, from which we can see that the daily index returns are skewed with excess kurtosis. As for the dynamic properties, the autocorrelations of index returns are in general low. For the squared return series, the first order autocorrelation is low but not negligible, and higher order autocorrelations are overall diminishing.

The GMM estimation, as proposed in section 3, is based on both the marginal and joint moment conditions of the observed return process Δs_t which is shown to be stationary.[6] There are obviously infinite number of moments that may be used in GMM estimation. The primary guidance of the moment selection in this paper is the Monte Carlo evidence in Andersen and Sørensen (1996) on the GMM estimation of a discrete-time SV model. Firstly, in determining the number of moments used in the estimation, we keep in mind the following fundamental trade-off: inclusion of additional moments improves estimation performance for a given degree of precision in the estimation of the weight matrix, but in finite sam-

[6]GMM is also used by Andersen (1994), Andersen and Sørensen (1996) to estimate discrete-time SV models and Ho, Perraudin and Sørensen (1996) to estimate a continuous-time SV model with different specification of the volatility process.

ples this must be balanced against the deterioration in the estimate of the weight matrix as the number of moments increases. Secondly, very high order moments should be avoided due to their erratic finite sample behavior caused by the presence of fat tails in the asset return distribution. Asymptotic normality of the GMM estimator requires finite variance of the moment conditions and good estimates of these quantities in finite samples. Thus our moment selection tends to focus on the lower order moments, which is consistent with Andersen and Sørensen (1996) and Jacquier, Polson and Rossi (1994). Thirdly, different from the discrete-time SV model, the absolute moments of the asset returns can not be derived for the continuous-time model. The Monte Carlo evidence in Andersen and Sørensen (1996), however, suggests that inclusion of these kinds of moments is in general unlikely to improve estimation performance and at best the gains are quite minor.

The exact moment conditions are chosen with further considerations of the particular model. First of all, since stochastic volatility allows for skewness and excess kurtosis, the first four unconditional moments are important. Secondly, since the autocorrelation of squared returns is determined by the SV process and its correlation with asset returns, the joint moments of the squared returns are important for the identification of the SV process. As the autocorrelation is varying over time, we use these moment conditions with different lags. Consequently, the moments included in the GMM estimation consist of the first four unconditional moments of the asset returns and the first five orders of autocorrelation of the squared returns. The weight matrix is estimated by the Barlett kernel proposed by Newey and West (1987) with a fixed lag length of 20. The starting values are set as the method of moments estimates of the parameters, which are obtained by matching the first four unconditional moments and the first order autocorrelation of the squared returns to the data.

The ECF estimation is also based on the time series observations of index returns Δs_t. We note that while the bivariate stochastic process $\{s_t, V_t\}$ is a Markov process by the continuous-time model specification, the marginal process $\{s_t\}$ and the return process $\{\Delta s_t\}$

are non-Markovian. Due to the lack of the exact optimal weight function in (20), our estimation is based on the minimization problem in (19) with the weight function being the pdf of a multivariate normal distribution with zero mean and variance-covariance matrix $\Sigma = \sigma_n^2 I$, namely $\varphi(u) = \frac{1}{\sqrt{(2\pi\sigma_n^2)^p}} \exp\{-\frac{u'u}{2\sigma_n^2}\}$. As we have pointed out in the general discussion of the ECF method, for non-Markovian processes there are no general solutions for the choice of optimal weight function $\varphi(\cdot)$ or the block size p. A practical solution for the choice of optimal weight function is to use the idea in the GMM procedure, i.e. choosing the weight function that minimizes the variance of the parameter estimates. In our particular choice of the weight function, it amounts to choosing the value of σ_n. The choice of the block size p is more complicated, depending on the exact structure of the DGP. For instance, for an AR(q) process we have that asymptotic efficiency will be achieved with p set to q (see Knight and Yu (2000)). However, for an invertible MA(q) process, which is in essence an AR(∞), it is clear that we need as large as possible block size to achieve asymptotic efficiency. For a general non-linear process, increasing the block size should in principle lead to gains in asymptotic efficiency, however, it also can substantially increase the computational burden in the estimation. Thus, in general there is a trade-off between large blocks for asymptotic efficiency and small blocks for computational efficiency. In addition, we note that with fixed sample size, increasing the block size will reduce the number of sampling blocks in the ECF estimation. In our application, to highlight the effect of varying block size and to ease the computational burden, we consider, in the ECF estimation, five different block sizes of two to six, i.e. $p = 1, 2, 3, 4$, and 5. With $p = 5$, the block size exceeds the span of one week period. The starting values of the parameters in the ECF estimation are also set as the method of moments estimates of the parameters.

Table 2 reports the parameter estimates and asymptotic standard errors for the SV model based on different estimation methods. For the ECF estimation, we only report for each p the parameter estimates using the multivariate normal pdf as weight function, with the standard deviation parameter chosen among various values to minimize the variance of parameter estimates. Firstly, the SV model has a reasonable fit to the S&P 500 index returns over the sample period. The Hansen-J χ^2 test statistic for the model specification based on GMM estimation has a p-value of 0.458.[7] Secondly, in both the GMM and ECF estimation, the estimate of β is in general significant, suggesting there exists mean reversion in the volatility process. Thirdly, the ECF estimates are in general sensitive to the choice of block size p. While the estimate of the expected asset return μ and that of the long run mean of volatility α (or $\sqrt{\alpha}$) are relatively stable, the estimates of parameters reflecting the dynamic properties of the model are, as expected, varying with different values of p and have relatively larger standard errors. In particular, the mean reversion parameter β has estimates that range from 0.214 to 0.313, and the asymmetric correlation parameter between the asset return and stochastic volatility has estimates that range from -0.244 to -0.273.

Many other estimation methods proposed for the dynamic latent variable model have also been applied to the SV model, mostly using the S&P 500 index returns as well. For instance, Andersen, Benzoni, and Lund (1998) and Chernov and Ghysels (1999) estimated the SV model using EMM based on daily S&P 500 index returns from 1980 to 1996 and from November 1985 to October 1994 respectively. Singleton (1999) estimated the SV model using SMM based on daily S&P 500 index returns over the period of 1990 to 1997.[8] Chacko and Viceira (1999) and Pan (1999) also estimated the SV model using weekly or monthly asset returns. In these studies, the data source, sampling period, or the drift function specifi-

[7] Based on the data set that includes the 1987 market crash, the SV model has rather poor fit to the S&P500 index returns and the GMM estimation reports a much smaller p-value for the Hansen-J χ^2 test statistic, suggesting that it is difficult for the SV model to generate large random jumps as in the case of 1987 market crash. Inclusion of the fifth moment of asset returns in the GMM estimation also significantly decreases the p-value of the Hansen-J χ^2 test statistic, which is a result of either the erratic finite sample behaviour caused by the presence of fat tails in the asset return distribution or the inability of the SV model to match the fifth moment of the observed process or the combination of both.

[8] In the simulation of sampling path, 50 subintervals per day are used in simulating 50,000 daily observations following the Euler discretization scheme.

cation varies from one to another. Overall, our estimation results based on the daily S&P500 index returns over the period of 1990 to 1999 are very comparable to the SMM estimates in Singleton (1999). Noticeable difference between our estimation results and those obtained by EMM is the value of the mean reversion parameter and the conditional variance of the volatility process. Our estimation results indicate a much stronger mean reversion and higher volatility for the stochastic volatility process. An intuitive justification is that, in the SV model framework, in order to incorporate the negative skewness and fat tails of the S&P 500 index return distribution, a negative correlation between asset return and volatility and a significant level of variation in the stochastic volatility are required. It is also noted that the parameters in the SV process are tied to each other as the unconditional mean and variance of the volatility are given respectively by α and $\alpha\frac{\sigma^2}{2\beta}$. It is easy to see that given α (the long-run mean of volatility) and the variance of the volatility, a high value of σ will induce a high value of β and *vice versa*.

A systematic comparison of the finite sample properties of alternative estimators, including the ones proposed in this paper, calls for a Monte Carlo simulation study. Due to the intensive computation involved in some of the estimation methods, especially the EMM approach, this task will not be pursued in this paper but in a separate study. To gauge the performance of our estimators, however, we investigate the implications of the parameter estimates on the static and dynamic properties of the underlying process. Intuitively, when the basic statistical properties of the estimated model are not consistent with those of the data, the following conclusions are in order: either the model is a poor candidate of the DGP or the estimation method employed has less desirable finite sample properties. Based on this intuition, we exploit the analytical moments derived from the model to calculate the major summary statistics from alternative parameter estimates. Table 3 presents the exact unconditional moments calculated from various parameter estimates, together with those calculated from the data. Overall, the GMM and ECF estimates provide the skewness and kurtosis of asset returns and the standard deviation of squared returns that are very close to their data counterparts. As expected, the increase of block size p does not significantly improve, if not hurt, the model's ability to capture the static properties of the data. Figure 1 further plots the first twenty orders of autocorrelations of the squared returns calculated from various parameter estimates, together with those calculated from the data. The GMM and ECF estimates all generate a relatively high first order autocorrelation with higher orders of autocorrelations quickly vanishing. For the ECF estimates, as p increases from 1 to 5, it is expected that the model can better capture the dynamic property of the data. As shown in Figure 1, the first five orders of autocorrelation of squared asset returns calculated from the ECF estimates appear to fit those calculated from the data reasonably well. However, the structure of the correlogram of the squared asset returns generated from the model is not successful in matching its counterpart of the data. More specifically, the correlogram of squared asset returns generated by the model is strictly monotonically decreasing, but the correlogram of squared asset returns calculated from the data, while overall decreasing, is not monotonic.

5 Conclusion

Concerning the statistical properties and the estimation of AD and AJD models, this paper has made several contributions. Firstly, we have shown how the joint characteristic function can be derived for AD and AJD models and thus enabling various statistical properties of these processes to be examined. As an illustration, from the joint characteristic function of asset returns, we derived analytically the *exact* static and dynamic properties of the continuous-time square-root SV process. Secondly, using the joint characteristic function we propose alternative estimators, both GMM and ECF, for these Markov AD and AJD models. More importantly, we show how in the latent variable case these estimators are both computationally efficient and asymptotically relatively efficient. The estimation is based on *exact* moment conditions or *exact* joint characteristic function and requires neither discretization of the continuous-time process nor simulation of the sampling path. Finally, the illustrative examples, in particular the SV model, have demonstrated the usefulness and elegance of the proposed techniques. The empirical application of the SV model using the S&P 500 index returns demonstrates that both the ECF and GMM estimation procedures not only are easy to implement and computationally efficient, but also have nice finite sample properties.

Appendix

Proof of Lemma 1: Using the following iterative procedure, from (2) we have

$$\begin{aligned}
& \psi(u_1, u_2, ..., u_{p+1}; x_1, x_2, ..., x_{p+1}|x_0) \\
= \; & E[\exp\{i \sum_{k=1}^{p+1} u_k' x_k\}|x_0] \\
= \; & E[\exp\{i \sum_{k=1}^{p} u_k' x_k\} E[\exp\{i u_{p+1}' x_{p+1}\}|x_p]|x_0] \\
= \; & E[\exp\{i \sum_{k=1}^{p} u_k' x_k\} \exp\{C(1, u_{p+1}) + D(1, u_{p+1})' x_p\}|x_0] \\
= \; & \exp\{C(1, u_{p+1})\} E[\exp\{i \sum_{k=1}^{p-1} u_k' x_k\} E[\exp\{i(u_p - iD(1, u_{p+1}))' x_p\}|x_{p-1}]|x_0]
\end{aligned}$$

Introducing the following notation,
$u_{p+1}^* = u_{p+1}$, and
$u_k^* = u_k - iD(1, u_{k+1}^*)$ for $k = 1, 2, ..., p$,
we have

$$\begin{aligned}
& \psi(u_1, u_2, ..., u_{p+1}; x_1, x_2, ..., x_{p+1}|x_0) \\
= \; & \exp\{C(1, u_{p+1}^*))E[\exp(i \sum_{k=1}^{p-1} u_k' x_k\} \exp\{C(1, u_p^*) + D(1, u_p^*)' x_{p-1}\}|x_0] \\
& ... \\
= \; & \exp\{\sum_{k=1}^{p+1} C(1, u_k^*)\} \exp\{D(1, u_1^*)' x_0\}
\end{aligned}$$

Taking expectations with respect to x_0 results in the joint unconditional characteristic function given by (5).

Proof of Lemma 2: Using the following iterative procedure, from (8) we have

$$\begin{aligned}
& \psi(u1_1, u1_2, ..., u1_{p+1}; s_1, s_2, ..., s_{p+1}|x_0) \\
= \; & E[\exp\{i \sum_{k=1}^{p+1} u1_k' s_k\}|x_0] \\
= \; & E[\exp\{i \sum_{k=1}^{p} u1_k' s_k\} E[\exp\{i u1_{p+1}' s_{p+1}\}|s_p, v_p]|x_0] \\
= \; & E[\exp\{i \sum_{k=1}^{p} u1_k' s_k\} \exp\{C(1, u1_{p+1}, 0) + D1(1, u1_{p+1}, 0)' s_p + D2(1, u1_{p+1}, 0)' v_p\}|x_0] \\
= \; & \exp\{C(1, u1_{p+1}, 0)\} E[\exp\{i \sum_{k=1}^{p-1} u1_k' s_k\} \times \\
& E[\exp\{i(u1_p - iD1(1, u1_{p+1}, 0))' s_p + D2(1, u1_{p+1}, 0))' v_p\}|s_{p-1}, v_{p-1}]|x_0]
\end{aligned}$$

Introducing the following notation:

$u1^*_p = u1_p, \quad u2^*_p = 0$, and

$u1^*_k = u1_k - iD1(1, u1^*_{k+1}, u2^*_{k+1}), \ u2^*_k = -iD2(1, u1^*_{k+1}, u2^*_{k+1})$ for $k = 1, 2, ..., p$,

we have

$$\begin{aligned}
& \psi(u1_1, u1_2, ..., u1_{p+1}; s_1, s_2, ..., s_{p+1}|x_0) \\
= \ & \exp\{C(1, u1^*_{p+1}, u2^*_{p+1})\}E[\exp\{i\sum_{k=1}^{p-1} u1'_k s_k\} \times \\
& \exp\{C(1, u1^*_p, u2^*_p) + D1(1, u1^*_p, u2^*_p)' s_{p-1} + D2(1, u1^*_p, u2^*_p)' v_{p-1}\}|x_0] \\
& \ldots \\
= \ & \exp\{\sum_{k=1}^{p+1} C(1, u1^*_k, u2^*_k\}) \exp\{D1(1, u1^*_1, u2^*_1)' s_0 + D2(1, u1^*_1, u2^*_1)' v_0\}
\end{aligned}$$

Taking expectations with respect to x_0 results in the unconditional joint characteristic function given in (9).

Proof of Lemma 3: Applying the results of Lemma 2 directly to the SV model, we derive the joint characteristic function of $\Delta s_1 = s_1 - s_0, \Delta s_2 = s_2 - s_1, ...,$ and $\Delta s_{p+1} = s_{p+1} - s_p$, where $p \geq 1$, i.e.,

$$\begin{aligned}
& \psi(u_1, u_2, ..., u_{p+1}; \Delta s_1, \Delta s_2, ..., \Delta s_{p+1}) \\
= \ & E[\exp\{iu_1 \Delta s_1 + iu_2 \Delta s_2 + ... + iu_{p+1} \Delta s_{p+1}\}] \\
= \ & E[E[\exp\{iu_1 \Delta s_1 + iu_2 \Delta s_2 + .. + iu_{p+1} \Delta s_{p+1}\} \mid s_p, v_p]] \\
= \ & E[\exp\{iu_1 \Delta s_1 + iu_2 \Delta s_2 + .. + iu_p \Delta s_p + C(1; u_{p+1}, 0) + D(1; u_{p+1}, 0) v_p\}]
\end{aligned}$$

Introducing the following notation,

$u1^*_{p+1} = u_{p+1}, u2^*_{p+1} = 0$

$u1^*_k = u_k, u2^*_k = -iD(1; u1^*_{k+1}, u2^*_{k+1})$ for $k = 1, ..., p$,

we have

$$\begin{aligned}
& \psi(u_1, u_2, ..., u_{p+1}; \Delta s_1, \Delta s_2, ..., \Delta s_{p+1}) \\
= \ & \exp\{C(1; u1^*_{p+1}, u2^*_{p+1})\}E[\exp\{i\sum_{k=1}^{p} u_k \Delta s_k + D(1; u1^*_{p+1}, u2^*_{p+1}) v_p\}] \\
& \ldots \\
= \ & \exp\{\sum_{k=1}^{p+1} C(1; u1^*_k, u2^*_k)\}E[\exp\{(\sum_{k=1}^{p+1} D(1; u1^*_k, u2^*_k)) v_0\}]
\end{aligned}$$

Taking expectations with respect to v_0 which follows a Gamma distribution, we have

$$\begin{aligned}
& \psi(u_1, u_2, ..., u_{p+1}; \Delta s_1, \Delta s_2, ..., \Delta s_{p+1}) \\
= \ & \exp\{\sum_{k=1}^{p+1} C(1; u1^*_k, u2^*_k)\} \times (1 - \frac{\sigma^2}{2\beta}\sum_{k=1}^{p+1} D(1; u1^*_k, u2^*_k))^{-2\beta\alpha/\sigma^2}
\end{aligned}$$

Proof of Lemma 4: We present details of how we calculate the asymptotic covariance matrix of the ECF estimator. Let our ECF estimator be given by $\hat{\theta}$ where

$$\hat{\theta} = \arg\min s(\theta)$$

and

$$\begin{aligned} s(\theta) &= \int \ldots \int \{(\text{Re}[\psi_n(u)] - \text{Re}[\psi(u,\theta)])^2 \\ &\quad + (\text{Im}[\psi_n(u)] - \text{Im}[\psi(u,\theta)])^2\} d\varphi(u) \end{aligned}$$

Now since

$$\text{Re}[\psi_n(u)] = \frac{1}{n}\sum_{j=1}^{n} \cos u' y_j$$

and

$$\text{Im}[\psi_n(u)] = \frac{1}{n}\sum_{j=1}^{n} \sin u' y_j$$

Then

$$\frac{\partial s(\theta)}{\partial \theta} = -\frac{2}{n}\sum_{j=1}^{n} K_j(\theta)$$

where

$$\begin{aligned} K_j(\theta) &= \left[\int \ldots \int \frac{\partial \text{Re}[\psi(u,\theta)]}{\partial \theta}(\cos u' y_j - \text{Re}[\psi(u,\theta)]) \right. \\ &\quad \left. + \frac{\partial \text{Im}[\psi(u,\theta)]}{\partial \theta}(\sin u' y_j - \text{Im}[\psi(u,\theta)])\right] d\varphi(u) \end{aligned}$$

Consequently

$$\sqrt{n}\frac{\partial s(\theta)}{\partial \theta} \xrightarrow{d} N(0, 4A(\theta))$$

where

$$A(\theta) = \lim_{n\to\infty} E\left[\frac{1}{n}\sum_j \sum_k K_j(\theta)K_k(\theta)\right]$$

and is given by

$$\begin{aligned} A(\theta) &= \lim_{n\to\infty} \frac{1}{n}\left\{\int \ldots \int \left[\frac{\partial \text{Re}[\psi(r,\theta)]}{\partial \theta}\frac{\partial \text{Re}[\psi(u,\theta)]}{\partial \theta}\sum_j \sum_k cov(\cos r' y_j, \cos u' y_k) \right.\right. \\ &\quad + \frac{\partial \text{Re}[\psi(r,\theta)]}{\partial \theta}\frac{\partial \text{Im}[\psi(u,\theta)]}{\partial \theta'}\sum_j \sum_k cov(\cos r' y_j, \sin u' y_k) \\ &\quad + \frac{\partial \text{Im}[\psi(r,\theta)]}{\partial \theta}\frac{\partial \text{Re}[\psi(u,\theta)]}{\partial \theta'}\sum_j \sum_k cov(\sin r' y_j, \cos u' y_k) \\ &\quad \left.\left. + \frac{\partial \text{Im}[\psi(r,\theta)]}{\partial \theta}\frac{\partial \text{Im}[\psi(u,\theta)]}{\partial \theta'}\sum_j \sum_k cov(\sin r' y_j, \sin u' y_k)\right]\right\} d\varphi(r) d\varphi(u) \end{aligned}$$

The double summation covariance expressions are readily found and given in the Lemma in Knight and Satchell [1997, p. 170]. That is, we note that

$$\begin{aligned} & \sum_j \sum_k cov(\cos r' y_j, \cos u' y_k) \\ = \; & n^2 cov(\text{Re}[\psi_n(r)], \text{Re}[\psi_n(u)]) \\ = \; & n^2 \cdot (\Omega_{RR})_{r,u} \end{aligned}$$

using notation in Knight and Satchell [1997]. Similarly, for the other double sums.

Thus

$$\begin{aligned} A(\theta) = & \lim_{n\to\infty} n \left\{ \int \ldots \int \left[\frac{\partial \text{Re}[\psi(r,\theta)]}{\partial \theta} \frac{\partial \text{Re}[\psi(u,\theta)]}{\partial \theta'} \cdot (\Omega_{RR})_{r,u} \right.\right. \\ & + 2 \frac{\partial \text{Re}[\psi(r,\theta)]}{\partial \theta} \frac{\partial \text{Im}[\psi(u,\theta)]}{\partial \theta'} \cdot (\Omega_{RI})_{r,u} \\ & \left.\left. + \frac{\partial \text{Im}[\psi(r,\theta)]}{\partial \theta} \frac{\partial \text{Im}[\psi(u,\theta)]}{\partial \theta'} \cdot (\Omega_{II})_{r,u} \right] \right\} d\varphi(r) d\varphi(u) \end{aligned}$$

Also we note that

$$\begin{aligned} E\left[\frac{\partial^2 s(\theta)}{\partial\theta\partial\theta'}\right] & = -\frac{2}{n}\sum_{j=1}^{n} E\left[\frac{\partial K_j(\theta)}{\partial\theta}\right] \\ & = \frac{2}{n}\sum_{j=1}^{n} \int \cdots \int \left[\frac{\partial \text{Re}[\psi(r,\theta)]}{\partial \theta} \frac{\partial \text{Re}[\psi(r,\theta)]}{\partial \theta'} \right. \\ & \quad \left. + \frac{\partial \text{Im}[\psi(r,\theta)]}{\partial \theta} \frac{\partial \text{Im}[\psi(r,\theta)]}{\partial \theta'} \right] d\varphi(r) \\ & = -2 \int \cdots \int \left[\frac{\partial \text{Re}[\psi(r,\theta)]}{\partial \theta} \frac{\partial \text{Re}[\psi(r,\theta)]}{\partial \theta'} \right. \\ & \quad \left. + \frac{\partial \text{Im}[\psi(r,\theta)]}{\partial \theta} \frac{\partial \text{Im}[\psi(r,\theta)]}{\partial \theta'} \right] d\varphi(r) \\ & = -2B(\theta) \end{aligned}$$

Thus standard asymptotic theory results in

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, B^{-1}(\theta) A(\theta) B^{-1}(\theta)).$$

Proof of Lemma 5: When

$$\ln L = \ln f(x_1) + \sum_{j=2}^{T} \ln f(x_j | I_{j-1})$$

MLE is achieved when

$$\frac{\partial \ln L}{\partial \theta} = \frac{\partial \ln f(x_1)}{\partial \theta} + \sum_{j=2}^{T} \frac{\partial \ln f(x_j|I_{j-1})}{\partial \theta} = 0$$

Now, from the definition of $w_1(r,\theta)$ and $w_j(r,\theta|I_{j-1})$, we have

$$\frac{\partial \ln f(x_1)}{\partial \theta} = \int w_1(r,\theta) e^{irx_1} dr$$

and

$$\frac{\partial \ln f(x_j|I_{j-1})}{\partial \theta} = \int w_j(r,\theta|I_{j-1}) e^{irx_j} dr$$

Substituting into the above equation involving the score function, we have

$$\frac{\partial \ln L}{\partial \theta} = 0 = \int w_1(r,\theta) e^{irx_1} dr + \sum_{j=2}^{n} \int w_j(r,\theta|I_{j-1}) e^{irx_j} dr = 0$$

References

Andersen, T. G., L. Benzoni, and J. Lund (1998), "Estimating jump-diffusions for equity returns" [Working paper, Northwestern University].

Andersen, T. G. and J. Lund (1997), "Estimating continuous time stochastic volatility models of the short term interest rate", *Journal of Econometrics*, **77**, 343–377.

Bailey, W. and E. Stulz (1989), "The pricing of stock index options in a general equilibrium model", *Journal of Financial and Quantitative Analysis,* **24**, 1–12.

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

Ball, C.A. and A. Roma (1994), "Stochastic volatility option pricing", *Journal of Financial and Quantitative Analysis*, **29**, 589–607.

Chacko, G. and L. M. Viceira (1999), "Spectral GMM estimation of continuous-time processes", *Working paper*, Graduate School of Business Administration, Harvard University.

Chan, K.C., Karolyi, G.A., Longstaff, F.A. and A.B. Sanders (1992), "An empirical comparison of alternative models of short-term interest rate ", *Journal of Finance*, **47**(3), 1209–1227.

Chernov, M. and E. Ghysels (1999), "A study toward a unified approach to the joint estimation of objective and risk-neutral measures for the purposes of options valuation", forthcoming *Journal of Financial Economics*.

Cox, J. C., J.E. Ingersoll, and S. A. Ross (1985a), "An intertemporal general equilibrium model of asset prices", *Econometrica,* **53**, 363–384.

Csörgö, S. (1981), "Limit behavior of the empirical characteristic function", *The Annals of Probability,* **9**, 130–144.

Dai, Q. and K.J. Singleton. (1999), "Specification analysis of affine term structure models", forthcoming *Journal of Finance*.

Duffie, D. and R. Kan (1996), "A yield-factor model of interest rates", *Mathematical Finance*, **6**(4), 379–406.

Duffie, D., J. Pan and K.J. Singleton (1999), "Transform analysis and asset pricing for affine jump-diffusions", forthcoming *Econometrica*.

Ethier, S. and T. Kurtz (1986), *Markov Processes, Characterization and Convergence*, New York: John Wiley and Sons.

Feuerverger, A. (1990), "An efficiency result for the empirical characteristic function in stationary time-series models", *The Canadian Journal of Statistics*, **18**, 155–161.

Feuerverger, A. and R.A. Mureika (1977), "The empirical characteristic function and its applications", *The Annals of Statistics*, **5**, 88–97.

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

Gallant, A. R and G. E. Tauchen (1996), "Which moments to match?", *Econometric Theory*, **12**, 657–681.

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

Heston, S. I. (1993), "A closed form solution for options with stochastic volatility with applications to bond and currency options", *Review of Financial Studies*, **6**, 327–344.

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

Kim, S., N. G. Shephard, and S. Chib (1996), "Stochastic volatility: optimal likelihood inference and comparison with arch models" [Forthcoming *Review of Economic Studies*].

Knight, J. L. and S.E. Satchell (1997), "The cumulant generating function estimation method", *Econometric Theory*, **13**, 170–184.

Knight, J. L. and J. Yu (2000), "Empirical Characteristic Function Estimation in Time Series", Working paper, University of Western Ontario.

Newey, W.K. and K.D. West (1987), "A simple positive definite heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica*, **55**, 703–708.

Pan, J. (1999), "Integrated time series analysis of spot and option prices", *Working paper*, Graduate School of Business, Stanford University.

Sandman, G. and S. J. Koopman (1997), "Maximum likelihood estimation of stochastic volatility models" [Working paper, London School of Economics].

Scott, L. (1987), "Option pricing when the variance changes randomly: Theory, estimators, and applications", *Journal of Financial and Quantitative Analysis*, **22**, 419–438.

Singleton, K. J. (1999), "Estimation of affine asset pricing models using the empirical characteristic function" [Manuscript, Stanford University].

Wong, E. (1964), "The construction of a class of stationary Markov processes: ", in *Stochastic Processes in Mathematical Physics and Engineering: Proceedings of symposia in Applied Mathematics*, Vol. **16**.

Table 1: Summary Statistics of S&P 500 Index Returns

Sample period: 1990 - 1999

Static Properties

	N	Mean	Std. Dev.	Skewness	Kurtosis	Max	Min
$\Delta s_t \times 100$	2527	5.581 10^{-2}	8.853 10^{-1}	-0.346	8.30	4.989	-7.113

Dynamic Properties (autocorrelations)

	N	$\rho(1)$	$\rho(2)$	$\rho(3)$	$\rho(4)$	$\rho(5)$	$\rho(10)$	$\rho(20)$
$\Delta s_t \times 100$	2527	0.016	0.003	-0.057	-0.014	-0.013	0.047	0.018
$(\Delta s_t \times 100)^2$	2527	0.213	0.141	0.074	0.074	0.115	0.052	0.050

Table 2: Estimation Results of the SV Model based on Different Methods

Sample Period	Method	μ	$\sqrt{\alpha}$	β	σ	ρ
	GMM	0.056	0.867	0.269	0.774	-0.271
		(0.017)	(0.021)	(0.131)	(0.295)	(0.107)
	ECF (p=1)	0.056	0.885	0.230	0.820	-0.273
		(0.017)	(0.023)	(0.101)	(0.201)	(0.114)
	ECF (p=2)	0.055	0.886	0.297	0.942	-0.246
		(0.016)	(0.023)	(0.128)	(0.283)	(0.117)
1990-1999	ECF (p=3)	0.056	0.874	0.313	0.960	-0.263
		(0.017)	(0.022)	(0.118)	(0.261)	(0.101)
	ECF (p=4)	0.059	0.871	0.274	0.773	-0.244
		(0.017)	(0.021)	(0.102)	(0.177)	(0.109)
	ECF (p=5)	0.059	0.863	0.214	0.713	-0.265
		(0.017)	(0.020)	(0.075)	(0.175)	(0.096)

Note: The numbers in brackets are standard errors of the estimates. The GMM estimates are based on the first four unconditional moments and first five cross moments of squared returns. The Hansen-J $\chi^2(4) = 3.628$ with p-value=0.458. ECF estimation is based on the minimization of integrated MSE of the analytical CF and ECF using normal probability density function as weight function with zero mean and the value of standard error chosen among various values to minimize the variance of the parameter estimates. $p+1$ represents the block size used in the application.

Table 3: Comparison between Data Moments and Model Moments Calculated from Different Parameter Estimates

	Std. of Δs_t	Skewness of Δs_t	Kurtosis of Δs_t	Std. of $(\Delta s_t)^2$
Data	0.885 10^{-2}	-0.346	8.303	2.108 10^{-4}
GMM	0.867 10^{-2}	-0.332	7.173	1.868 10^{-4}
ECF (p=1)	0.885 10^{-2}	-0.352	8.307	2.117 10^{-4}
ECF (p=2)	0.886 10^{-2}	-0.356	8.302	2.121 10^{-4}
ECF (p=3)	0.874 10^{-2}	-0.391	8.366	2.073 10^{-4}
ECF (p=4)	0.871 10^{-2}	-0.297	7.026	1.862 10^{-4}
ECF (p=5)	0.863 10^{-2}	-0.306	7.547	1.906 10^{-4}

Note: The data moments of S&P 500 index returns are calculated from daily returns over the sample period from 1990 to 1999. Parameter estimates using GMM and ECF are based on the same data set.

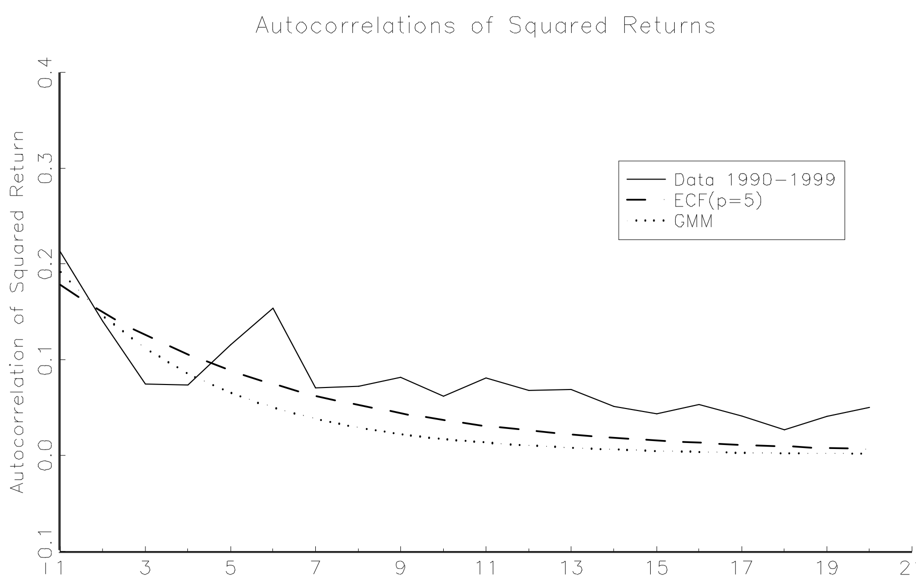

Figure 1: Correlogram of Squared Asset Return