Parameter Estimation and Model Testing for Markov Processes via Conditional Characteristic Functions

Song Xi Chen[a,b], Liang Peng[c] and Cindy L. Yu[a]

[a] Department of Statistics, Iowa State University, Ames, Iowa 50011-1210, USA

[b] Guanghua School of Management and Center for Statistical Science

Peking University, Beijing 100871, China

[c] School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332

Abstract: Markov processes are used in a wide range of disciplines including finance. The transition densities of these processes are often unknown. However, the conditional characteristic functions are more likely to be available especially for Lévy driven processes. We propose an empirical likelihood approach for both parameter estimation and model specification testing based on the conditional characteristic function for processes with either continuous or dis-continuous sample paths. Theoretical properties of the empirical likelihood estimator for parameters and a smoothed empirical likelihood ratio test for a parametric specification of the process are provided. Simulations and empirical case study are carried out to confirm the effectiveness of the proposed estimator and test.

Keyword: Conditional characteristic function; Diffusion processes; Empirical likelihood; Kernel smoothing; Lévy driven processes

1 Introduction

Let $\{X_t(\theta)\}_{t\in\mathcal{T}}$ be a parametric d-dimensional Markov process defined by

$$dX_t = \mu(X_t;\theta)dt + \sigma(X_t;\theta)dL_{t;\theta}, \tag{1.1}$$

where $\mu(\cdot)$ is a d-dimensional drift function, $\sigma(\cdot)$ is a $d \times d$ matrix-valued function of X_t, $L_{t;\theta}$ is a Lévy process in R^d, and $\theta \in \Theta \subset R^p$. When L_t is a standard Brownian motion, (1.1) is a diffusion process having a continuous sample path. When L_t contains the Brownian motion and a compound Poisson process, (1.1) becomes the jump diffusion process. A stochastic process of form (1.1) has long been used to model stochastic systems arising in physics, biology and other natural sciences. It has also been the fundamental tool in financial modeling. We refer to Sundaresan (2000) and Fan (2005) for overviews, Barndorff-Nielsen, Mikosch and Resnick (2001) for recent developments on Lévy driven processes, and Sørensen (1991) for statistical inference. Important subclasses of (1.1) include i) the multivariate diffusion process defined by

$$dX_t = \mu(X_t;\theta)dt + \sigma(X_t;\theta)dB_t, \tag{1.2}$$

where B_t is the standard Brownian motion in R^d (Stroock and Varadhan, 1979 and Øksendal, 2000); ii) the Vasicek with Merton Jump model (VSK-MJ) defined by

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dB_t + J_t dN_t, \tag{1.3}$$

where κ, α and σ are unknown parameters and represent the mean reverting rate, long-run mean and volatility of the process respectively, N_t is a Poisson process with intensity λ, and J_t is the random jump size independent of the filtration $\mathcal{F}_t$ up to time t and has a normal density $N(0, \eta^2)$ (Merton, 1976); iii) Lévy driven Ornstein-Uhlenbeck process defined by

$$dX_t = -\lambda X_t dt + dL_{\lambda t}, \qquad X_0 > 0, \tag{1.4}$$

where L_t is a Lévy process with no Brownian part, a non-negative drift and a Lévy measure which is zero on the negative half line, and the parameter λ is positive (see Barndorff-Nielsen and Shephard, 2001).

Often a closed form expression for the transition density of process (1.1) is not available except for some special processes, even the transition density exists and is unique. This fact prevents the use of the maximum likelihood estimation (MLE) and the specification tests based on the exact transition density. Recently Aït-Sahalia(2002, 2008) established expansions for the transition densities so that parameter estimation can be based on the approximate likelihood functions. Testing may be also formulated via the approximate density; see Chen, Gao and Tang (2008) and Aït-Sahalia, Fan and Peng (2009) for such tests. The conditional characteristic functions (CCF) are more likely available than the transition densities for the continuous-time models, especially for the Lévy driven processes through the celebrated Lévy-Khintchine representation. For instance, Duffie, Pan and Singleton (2000) derived the explicit form of the CCF for multivariate affine jump processes, which include the Vasicek with Merton jump process given in (1.3). The CCF for the Lévy driven Ornstein-Uhlenbeck process (1.4) is established in Barndorff-Nielsen and Shephard (2001).

Statistical inference based on the characteristic functions was proposed by Feuerverger and Mureika (1977), Feuerverger and McDunnough (1981) for independent observations and Feuerverger (1990) for discrete time series. Singleton (2001) introduced the approach to inference for parametric continuous-time Markov processes and show that estimation can be carried out based on the CCF without having to carry out the the Fourier inversion. Chacko and Viceira (2003) proposed a generalized method of moment estimator (GMM) for parameters at a finite number of frequencies of the CCF. Carrasco, Chernov, Florens and Ghysels (2007) carried out GMM estimation on a slowly diverging number of frequencies of the CCF to achieve the optimal estimation efficiency offered by the MLE. Jiang and Knight (2002) proposed GMM estimators based on the joint characteristic function of the observed state variables. Chen and Hong (2010) proposed a test for multivariate processes based on the CCF via a generalized spectral density approach.

In this paper, we first propose an empirical likelihood (Owen, 1988) approach for parameter estimation and model specification testing of a parametric Markov process via the CCF. An empirical likelihood ratio is formulated for the unknown parameters assuming the specification (1.1), which leads to a non-parametric maximum likelihood estimator. The proposed estimator may be viewed as a compromise between Chacko and Viceira (2003)'s GMM based on a finite number of frequencies and that of Carrasco, Chernov, Florens and Ghysels (2007) of a high dimensional GMM. The high dimensional GMM approach requires ridging a high dimensional weighting matrix in order to avoid its singularity, and the selecting the ridging parameter can be computationally expensive. The proposed estimation utilizes a wide range of frequency information in the parametric CCF, while having the computation easily managed.

We then formulate an empirical likelihood CCF based model specification test for the parametric process (1.1) via kernel smoothing. The proposed test extends the transition density based tests of Hong and Li (2005), Chen, Gao and Tang (2008) and Aït-Sahalia, Fan and Peng (2009) to the CCF based. This largely increases the range of the continuous-time Markov processes which can be tested directly without replying on the transition density approximation. The proposed test provides an alternative formulation of the CCF based test of Chen and Hong (2010), which is based on an explicit L_2 measure between an kernel estimator of the CCF and its parametric counter-part. It is largely distinct from the above mentioned tests, except Chen and Hong (2010), by targeting directly on CCF, which is more readily available for continuous-time models than the transition density functions. Another advantage of the proposed test is the empirical likelihood (EL) formulation, which can produce an integrated likelihood ratio test in a nonparametric setting. The proposed test utilizes some of the attractive properties of the EL, like internal studentizing without an explicit variance estimation and good power performance. How to extend the proposed methods to the case of latent variables is quite challenging and will be a part of our future research.

The paper is organized as follows. In Section 2, we introduce and evaluate the CCF based empirical likelihood estimator. The model specification test is given in Section 3. Section 4 reports results from simulation studies. An empirical study for a set of 3-month treasury bill rate data is analyzed in Section 5. All technical details are reported in the Appendix.

2 Parameter Estimation

Let $\{X_{t\delta}\}_{t=1}^{n}$ be n discretely sampled observations of (1.1). For notation simplification, we denote $X_{t\delta}$ as X_t, where the sampling interval δ is any fixed quantity. Let $\psi_t(u;\theta) = E_\theta(e^{iu^T X_{t+1}}|X_t)$, for $u \in R^d$, be the conditional characteristic function. We use $\bar{a}$ and $A^\star$ to denote the conjugate of a complex number a and the conjugate transpose of the complex matrix A, respectively.

Let $\epsilon_t(\tau;\theta) = w(u,r;X_t)\{e^{iu^T X_{t+1}} - \psi_t(u;\theta)\}$ for $\tau = (u^T, r^T)^T \in R^{2d}$, where $w(u,r;X_t)$ is a weight factor. Here $\epsilon_t(\tau;\theta)$ can be regarded as 'residuals' between $e^{iu^T X_{t+1}}$ and the parametric CCF $\psi_t(u;\theta)$. The complex weight factor $w(u,r;X_t)$ satisfies $\bar{w}(u,r;X_t) = w(-u,-r;X_t)$ and $|w(u,r;X_t)| = 1$ for any $u, r \in R^d$, whose use is aimed to utilize more model information. Let θ_0 be the true parameter and the unique solution of

$$E\{e^{iu^T X_{t+1}} - \psi_t(u;\theta)|X_t\} = 0 \quad \text{for all } u \in R^d. \tag{2.1}$$

From the Markov property and (2.1), for any $\tau = (u^T, r^T)^T \in R^{2d}$

$$E\{\epsilon_t(\tau;\theta_0)\} = 0 \quad \text{and} \quad Cov\{\epsilon_{t_1}(\tau;\theta_0), \epsilon_{t_2}(\tau;\theta_0)\} = 0 \quad \text{if} \quad t_1 \neq t_2. \tag{2.2}$$

Let $\epsilon_t^R(\tau;\theta)$ and $\epsilon_t^I(\tau;\theta)$ be the real and imaginary parts of $\epsilon_t(\tau;\theta)$ respectively, and $\vec{\epsilon}_t(\tau;\theta) = \left(\epsilon_t^R(\tau;\theta), \epsilon_t^I(\tau;\theta)\right)^T$ be the real bivariate vector corresponding to $\epsilon_t(\tau;\theta)$.

We now formulate an empirical likelihood for θ based on the CCF $\psi_t(u;\theta)$. The empirical likelihood (EL) introduced in Owen (1988) is a technique that allows construction of a non-parametric likelihood for parameters of interest. Despite that the EL method is intrinsically non-parametric, it possesses two important properties of a parametric likelihood, the Wilks' theorem and the Bartlett correction; see Chen and Van Keilegom (2009) for a latest overview and Kitamura, Tripathi and Ahn (2004) for a formulation with conditional moments.

Let $p_1(\tau), \cdots, p_n(\tau)$ be probability weights allocated to the 'residuals' $\{\vec{\epsilon}_t(\tau;\theta)\}_{t=1}^n$. A local EL for θ at τ is

$$L_n(\tau,\theta) = \max \prod_{t=1}^{n} p_t(\tau) \tag{2.3}$$

subject to $\sum_{t=1}^n p_t(\tau) = 1$ and $\sum_{t=1}^n p_t(\tau)\vec{\epsilon}_t(\tau;\theta) = 0$. Here the second constraint reflects (2.1). The maximum empirical likelihood is attained at $p_t(\tau) \equiv n^{-1}$ for all t such that the maximum likelihood $L_n(\tau;\theta) = n^{-n}$. Let $\ell_n(\tau;\theta) = -2\log\{L_n(\tau;\theta)/n^{-n}\}$ be the local log-EL ratio of θ at τ.

Employing the EL algorithm (Owen, 1988), the optimal $p_t(\tau)$ of the above optimization problem (2.3) is

$$p_t(\tau) = \frac{1}{n}\frac{1}{1+\lambda(\tau;\theta)^T\vec{\epsilon}_t(\tau;\theta)},$$

where $\lambda(\tau;\theta)$ is a Lagrange multiplier in R^2 that satisfies

$$Q_{1n}(\tau;\theta,\lambda) =: \frac{1}{n}\sum_{t=1}^{n}\frac{\vec{\epsilon}_t(\tau;\theta)}{1+\lambda(\tau;\theta)^T\vec{\epsilon}_t(\tau;\theta)} = 0. \tag{2.4}$$

Hence, the local EL ratio becomes

$$\ell_n(\tau;\theta) = 2\sum_{t=1}^{n}\log\{1+\lambda(\tau;\theta)^T\vec{\epsilon}_t(\tau;\theta)\}. \tag{2.5}$$

Integrating $\ell_n(\tau;\theta)$ against a probability weight $\pi(\tau)$ which is supported on a compact set S in R^{2d}, an integrated empirical likelihood ratio for θ is

$$\ell_n(\theta) = \int_{\tau\in R^{2d}} \ell_n(\tau;\theta)\pi(\tau)d\tau. \tag{2.6}$$

The maximum EL estimator (MELE) for θ is defined as

$$\hat{\theta}_n = \arg\min_{\theta} \ell_n(\theta),$$

by noting that -2 has been multiplied in the EL ratio $\ell_n(\tau;\theta)$.

Like Qin and Lawless (1994), we first show that there exists a consistent estimator $\hat{\theta}_n$ with a certain rate of convergence as follows.

Lemma 1. Under conditions C1-C4 given in the appendix, with probability one, $\ell_n(\theta)$ attains its minimum at $\hat{\theta}_n$ in the interior of the ball $||\theta-\theta_0|| \le O(n^{-1/3})$, and $\hat{\theta}_n$ and $\lambda(\tau;\hat{\theta}_n)$ satisfy

$$\begin{cases} Q_{1n}(\tau;\hat{\theta}_n,\lambda(\tau;\hat{\theta}_n)) = 0 \quad \text{for all} \quad \tau \in S \quad \text{and} \\ \int Q_{2n}(\tau;\hat{\theta}_n,\lambda(\tau;\hat{\theta}_n))\pi(\tau)\,d\tau = 0, \end{cases} \tag{2.7}$$

where Q_{1n} is defined in (2.4) and

$$Q_{2n}(\tau;\theta,\lambda) = \frac{1}{n}\sum_{t=1}^{n}\frac{1}{1+\lambda(\tau;\theta)^T\vec{\epsilon}_t(\tau;\theta)}\frac{\partial\vec{\epsilon}_t^{\,T}(\tau;\theta)}{\partial\theta}\lambda. \tag{2.8}$$

Before deriving the asymptotic normality of the $\hat{\theta}_n$, we define

$$M_0 = \tfrac{1}{2}\begin{pmatrix} 1 & 1 \\ i^{-1} & -i^{-1} \end{pmatrix}, \quad \tilde{\epsilon}_t(\tau;\theta) = (\epsilon_t(\tau;\theta),\epsilon_t(-\tau;\theta))^T,$$

$A(\tau_1,\tau_2;\theta_0,\theta) = Cov\{\tilde{\epsilon}_1(\tau_1;\theta),\tilde{\epsilon}_1(\tau_2;\theta)\}$,

$$\Gamma(\theta_0) =: \int E\left(\frac{\partial\tilde{\epsilon}_1^{\star}(\tau;\theta_0)}{\partial\theta}\right)A^{-1}(\tau,\tau;\theta_0,\theta_0)E\left(\frac{\partial\tilde{\epsilon}_1(\tau;\theta_0)}{\partial\theta}\right)\pi(\tau)d\tau, \tag{2.9}$$

and

$$\begin{aligned} V(\theta_0) &= \int\int E\left(\tfrac{\partial\tilde{\epsilon}_1^{\star}(\tau_1;\theta_0)}{\partial\theta}\right)A^{-1}(\tau_1,\tau_1;\theta_0,\theta_0)A(\tau_1,\tau_2;\theta_0,\theta_0) \\ &\quad \times A^{\star-1}(\tau_2,\tau_2;\theta_0,\theta_0)E\left(\tfrac{\partial\tilde{\epsilon}_1(\tau_2;\theta_0)}{\partial\theta}\right)\pi(\tau_1)\pi(\tau_2)d\tau_1 d\tau_2. \end{aligned} \tag{2.10}$$

Theorem 1: Under Conditions C1-C4 given in the appendix, for the estimator $\hat{\theta}_n$ in Lemma 1, we have $\sqrt{n}(\hat{\theta}_n-\theta_0) \xrightarrow{d} N(0,\Sigma)$ where $\Sigma = \Gamma^{-1}(\theta_0)V(\theta_0)\Gamma^{-1}(\theta_0)$.

The proposed estimator attains the $\sqrt{n}$-rate of convergence. It is computationally stable because computing $\ell_n(\tau;\theta)$ for one τ at a time is essentially one dimensional problem. Note that Carrasco, Chernov, Florens and Ghysels (2007) considered CCF based generalized method of moment estimation by considering a continuum of τ's in a functional space via covariance operator, but the covariance operator may not be invertible due to zero eigenvalues. Hence, Carrasco, Cherno, Florens and Ghysels (2007) needed ridging to avoid the invertible issue, which makes the computation quite involved. When the dimension of θ is not small, it would be useful to replace the equation $\sum_{t=1}^{n} p_t \vec{\epsilon}_t(\tau;\theta) = 0$ in formulating the local EL in (2.3) by several equations such as

$$\sum_{t=1}^{n} p_t \vec{\epsilon}_t(s_1\tau;\theta) = 0, \cdots, \sum_{t=1}^{n} p_t \vec{\epsilon}_t(s_m\tau;\theta) = 0$$

for some given $s_1, \cdots, s_m$.

3 Test for Model Specification

In this section we consider testing for the validity of (1.1) via testing for the parametric specification of the CCF $\psi_t(u;\theta)$. Tests for model specification of a continuous-time Markov process have been proposed by Chen, Gao and Tang (2008) and Aït-Sahalia, Fan and Peng (2009). Despite parameter estimation based on the transition density is asymptotically efficient, it is unclear if a test based on the transition density is more powerful than one based on the CCF. The choice is clearer when the transition density does not admit a closed form while the CCF does, since the latter is a test valid at any level of the sampling interval δ.

Let the underlying process that generates the observed sample path $\{X_t\}_{t=1}^{n}$ be

$$dX_t = \mu(X_t)dt + \sigma(X_t)dL_t, \tag{3.1}$$

whose CCF is $\psi(u;X_t)$. The process (1.1) is a parametric specification of (3.1). To emphasize the dependence of the CCF on X_t, we write in this section $\psi_t(u)$ as $\psi(u, X_t)$, $\psi_t(u;\theta)$ as $\psi(u, X_t;\theta)$ and other quantities in a similar fashion. We consider testing

$$H_0 : P\{\psi_t(u) = \psi_t(u;\theta_0)\} = 1 \quad \text{for all } u \in R^d \text{ and some } \theta_0 \in \Theta,$$

against a sequence of local alternative hypotheses

$$H_1 : P\{\psi_t(u) = \psi_t(u;\theta_0) + c_n\Delta_n(u;X_t)\} = 1 \quad \text{for all } u \in R^d,$$

where $\{c_n\}$ is a sequence of non-random real constants converging to zero at a certain rate, and $\{\Delta_n(u;X_t)\}$ is a sequence of bounded complex functions which are continuous at $u = 0$ and $\Delta_n(0;X_t) \equiv 0$; see Condition C6 in the appendix for extra restrictions.

Since the target of inference is a conditional quantity, we need to work with a kernel smoothed version of $\ell_n(\theta)$. Let K be a kernel function which is a symmetric probability density in R^d, and h be a smoothing bandwidth that tends to 0 as $n \to \infty$. A smoothed version of $L_n(\tau,\theta)$ is

$$L_{nh}(\tau, x;\theta) = \max \prod_{t=1}^{n} p_t(\tau, x) \tag{3.2}$$

subject to $\sum_{t=1}^{n} p_t(\tau, x) = 1$ and $\sum_{t=1}^{n} p_t(\tau, x)K_h(x - X_t)\vec{\epsilon}(\tau, X_t;\theta) = 0$.

Let $\ell_{nh}(\tau, x, \theta) = -2\log\{L_{nh}(\tau, x, \theta)n^n\}$ be the log-EL ratio. Then, the integrated log-EL ratio for θ is

$$\ell_{nh}(\theta) = \int\int \ell_{nh}(\tau, x, \theta)\pi_1(\tau)\pi_2(x)d\tau dx$$

where π_1 and π_2 are probability weight functions on the frequency space and the state space respectively. We can choose π_1 to be the same as the π in Section 3.

The test statistic is $\ell_{nh}(\hat{\theta}_n)$, where $\hat{\theta}_n$ is the empirical likelihood estimator proposed in Section 3. As a matter of fact, we can employ any estimator with $n^{1/2}$-rate of convergence. To appreciate the meaning of the test statistic, let $W_h(x-X_t) = K_h(x-X_t)/\sum_{j=1}^{n} K_h(x-X_j)$ be the Nadaraya-Watson kernel weight, $\epsilon_{n,h}(\tau, x;\theta) = \sum_{t=1}^{n} K_h(x - X_t)\epsilon(\tau, X_t;\theta)$ be the kernel smooth of the residuals,

$\tilde{\epsilon}_{n,h}(\tau, x;\theta) = (\epsilon_{nh}(\tau, x;\theta), \epsilon_{nh}(-\tau, x;\theta))^T$, and $R(K) = \int K^2(t)dt$. It can be shown by a similar derivation in Chen, Härdle and Li (2003) that

$$\begin{aligned} \ell_{nh}(\theta) \;=\; & nh^d R^{-1}(K) \int\int \tilde{\epsilon}^{\star}_{n,h}(\tau, x;\theta)V^{-1}(\tau, x;\theta_0, \theta)\tilde{\epsilon}_{n,h}(\tau, x;\theta) \\ & \times\pi_1(\tau)f(x)\pi_2(x)d\tau dx + O_p\{(nh^d)^{-1/2}\log^3(n) + h^2\log^2(n)\}, \end{aligned} \tag{3.3}$$

where $V(\tau, x; \theta_0, \theta) = Var\{\tilde{\epsilon}(\tau, X_t; \theta)|X_t = x\}$ and $f(x)$ is the density of X_t. So, the test statistic is asymptotic equivalent to a L_2-measure of the averaged 'residuals' $\tilde{\epsilon}^{\star}_{n,h}(\tau, x; \theta)$ inversely weighted by the covariance matrix function V. Hence, the proposed test is similar in tune to Fan and Zhang (2003) for testing diffusion processes, and of Härdle and Mammen (1993) and Wang and Van Keilegom (2007) for testing regression functions.

We need the following notations to describe the power property. Let $V(\tau_1, \tau_2, x) = E\{\tilde{\epsilon}(\tau_1, X_t; \theta_0)\tilde{\epsilon}^{\star}(\tau_2, X_t; \theta_0)|X_t = x\}$, then $V(\tau, \tau, x; \theta_0, \theta_0) = V(\tau, x)$ defined earlier. Express the matrices

$$V(\tau_1, \tau_2, x) = (V_{lk}(\tau_1, \tau_2, x))_{1\le l,k\le 2} \quad \text{and} \quad V^{-1}(\tau, x) = \left(\nu^{lk}(\tau, x)\right)_{1\le l,k\le 2}.$$

Furthermore, we choose $c_n = n^{-1/2}h^{d/4}$ and define

$$\eta_n(\tau, X_t) = w(\tau; X_t)\Delta_n(u, X_t), \quad \tilde{\eta}_n(\tau, X_t) = (\eta_n(\tau, X_t), \eta_n(-\tau, X_t))^T,$$

$$\mu_n = \int\int \tilde{\eta}^{\star}_n(\tau, x)V^{-1}(\tau, x; \theta_0, \theta_0)\tilde{\eta}_n(\tau, x)\pi_1(\tau)\pi_2(x)f(x)d\tau dx,$$

$\sigma_n^2 = 2R^{-2}(K)h^{-d}\gamma^2(K, V, \pi_1, \pi_2)$ where

$$\begin{aligned}\gamma^2(K, V, \pi_1, \pi_2) \;&=\; K^{(4)}(0)\int\int\int \sum_{l_1,k_1,l_2,k_2}^{2} V_{l_1 l_2}(-\tau_1, \tau_2, x)V_{k_1 k_2}(\tau_1, -\tau_2, x)\nu^{l_1,k_1}(\tau_1, x)\\ &\times\;\; \nu^{l_2,k_2}(\tau_2, x)\pi_1(\tau_1)\pi_1(\tau_2)\pi_2^2(x)d\tau_1 d\tau_2 dx. \qquad (3.4)\end{aligned}$$

where $K^{(4)}$ is the 4-th convolution of the kernel function K.

The asymptotic normality of $\ell_{nh}(\hat{\theta}_n)$ is given in the following theorem.

Theorem 2 Under Conditions C1-C6 given in the appendix,

$$h^{-d/2}(\ell_{nh}(\hat{\theta}_n) - 2 - h^{d/2}\mu_n) \xrightarrow{d} N(0, 2R^{-2}(K)\gamma^2(K, V, \pi_1, \pi_2)). \qquad (3.5)$$

We note that $\mu_n = 2$ under H_0. Under H_1, since $\Delta_n(u, x)$ is non-vanishing with respect to u, $\tilde{\eta}_n(\tau, x)$ is non-vanishing with respect to u for all x in the support of f, which leads to a positive quantity μ_n due to $V^{-1}(\tau, x; \theta_0, \theta_0)$ being a Hermitian matrix. Since no restriction has been imposed on the functional form of $\Delta_n(u, X_t)$, it means that the test is powerful for a wide range of local alternatives. Indeed, if $\hat{\gamma}^2(K, V, \pi_1, \pi_2)$ is a consistent estimator of $\gamma^2(K, V, \pi_1, \pi_2)$, the asymptotic normality based test for H_0 with α-level of significance rejects H_0 if

$$\ell_{nh}(\hat{\theta}_n) \geq 2 + z_{1-\alpha}\sqrt{2}h^{d/2}R^{-1}(K)\hat{\gamma}(K, V, \pi_1, \pi_2),$$

where $z_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution. Theorem 2 implies that the power of the test under H_1 is

$$\Phi\left(-z_{1-\alpha} + \frac{R(K)\mu_n}{\sqrt{2}\gamma(K, V, \pi_1, \pi_2)}\right),$$

where Φ is the standard normal distribution function.

It is known that the choice of bandwidth is important in any test based on the kernel smoothing technique. To make the test less sensitive to the choice of smoothing bandwidth, we propose carrying out the test based on a set of bandwidths, say $\{h_1, \cdots, h_k\}$, for a fixed integer k such that $h_i = c_i h$ for some constants $c_1 < c_2 < ... < c_k$. Here h is a reference bandwidth which may be obtained via the cross-validation method.

This means that we have a set of the EL ratios $\{\ell_{nh_1}(\hat{\theta}_n), \cdots, \ell_{nh_k}(\hat{\theta}_n)\}$ corresponding to the bandwidth set, and the overall test statistic is

$$T_n = \max_{1 \leq i \leq k}\{h_i^{-d/2}(\ell_{nh_i}(\hat{\theta}_n) - 2)\}. \tag{3.6}$$

To describe the asymptotic distribution of T_n, let $K^{(2)}(z, c) = \int K(u)K(z + cu)du$ be a generalization to the convolution of K, $\nu(t) = \int \{K^{(2)}\,(tu, t)\}^2 du$ and

$$\Sigma_J = \frac{2}{R^2(K)} \int\int \pi_1(\tau_1)\pi_1(\tau_2)\pi_2^2(x)dxd\tau_1 d\tau_2 \left((c_j/c_i)^d \nu(c_i/c_j)\right)_{J\times J}.$$

Theorem 3. Under Conditions C1-C6, $T_n \xrightarrow{d} \max_{1\leq k\leq J} Z_k$ as $n \to \infty$ where

$$(Z_1, \ldots, Z_J)^T \sim N(0, \Sigma_J).$$

Let t_α be the $1-\alpha$ quantile of T_n where $\alpha \in (0,1)$ is the nominal size of the test. The following parametric bootstrap procedure is employed to approximate t_α:

Step 1: Simulate a sample path $\{X_t^*\}_{t=1}^n$ at the same frequency δ according to the model under H_0 with the CCF based estimate $\hat{\theta}_n$.

Step 2: Let $\tilde{\theta}_n^*$ be the estimate of θ under H_0 using the resample path $\{X_t^*\}_{t=1}^n$ obtained in Step 1, and T_n^* be the version of T_n for the resampled path.

Step 3: For a large positive integer B, repeat Steps 1 and 2 B times and obtain, after ranking, $T_n^{(1)*} \le T_n^{(2)*} \le \cdots \le T_n^{(B)*}$.

Then, the Monte Carlo approximation of t_α is $T_n^{([B(1-\alpha)]+1)*}$. The proposed test rejects H_0 if $T_n(\hat{\theta}_n) \ge T_n^{([B(1-\alpha)]+1)*}$. The justification of the above bootstrap procedure can be made based on Theorem 3 via the standard techniques for instance those given in Chen, Gao and Tang (2008).

4 Simulation Study

We report in this section the results from our simulation studies which are designed to verify the proposed parameter estimator and model testing procedure. To evaluate the quality of the proposed EL estimator, we first chose two univariate diffusion processes with known transition densities, so that the MLEs can be compared with the proposed EL estimates. The two processes are the Vasicek model (Vasicek, 1977) (VSK),

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dB_t, \tag{4.1}$$

and the Cox-Ingersoll-Ross Model (Cox, Ingersoll and Ross, 1985) (CIR),

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dB_t, \tag{4.2}$$

where κ, α and σ are unknown parameters which represent the mean reverting rate, long-run mean and volatility of the process respectively. Both processes are widely used in interest rate modeling and various option price formulation. For the Vasicek model, the transition distribution of $X_{t+1}|X_t$ is a normal distribution $N(\alpha+(X_t-\alpha)exp(-\kappa\delta), \sigma^2(1-exp(-2\kappa\delta))/(2\kappa))$. For the CIR model, when $2\kappa\alpha/\sigma^2 > 1$, $X_{t+1}|X_t$ is a multiple of a non-central Chi-square random variable with degrees of freedom $4\kappa\alpha/\sigma^2$ and non-centrality parameter $cX_t exp(-\kappa\delta)$, where the multiplier is $1/c$ with $c = 4\kappa/(\sigma^2(1-exp(-\kappa\delta)))$. The CCFs of these two models can easily be derived from their known transitional densities.

We then considered estimation for the jump diffusion model VSK-MJ as given in (1.3) based on its CCF function

$$\psi_t(u;\theta) = exp\{\frac{\sigma^2 u^2}{4\kappa}(e^{-2\kappa\delta} - 1) - \lambda\delta + \gamma + i(\alpha u(1 - e^{-\kappa\delta}) + ue^{-\kappa\delta}X_t)\}, \tag{4.3}$$

where $\gamma = \lambda/(2\kappa)\int_{e^{-2\kappa\delta}}^{1} exp(-\eta^2 u^2 y/2)/y dy$. For comparison, we approximated its transition density by a mixture of normal distributions, $(1-\lambda\delta)N(\mu_\delta, \sigma_\delta^2) + \lambda\delta N(\mu_\delta, \sigma_\delta^2 + \eta^2)$, which is a first order approximation proposed in Aït-Sahalia, Fan and Peng (2009). Here, $\mu_\delta = \alpha + (X_t - \alpha)exp(-\kappa\delta)$, and $\sigma_\delta^2 = \sigma^2(1 - exp(-2\kappa\delta))/(2\kappa)$. The approximate MLEs were obtained based on the mixture approximation given above.

We also consider the Inverse Gaussian OU process (IG-OU) in (1.4), i.e. the process X_t follows the Inverse Gaussian law $IG(a, b)$, for every t when X_0 is generated from $IG(a, b)$. The CCF of this process is

$$\psi_t(u;\theta) = exp\{-a(\sqrt{-2iu + b^2} - \sqrt{-2iue^{-\lambda\delta} + b^2}) + iue^{-\lambda\delta}X_t\}. \tag{4.4}$$

Since neither the exact transition density nor its approximation is available, we were content with carrying out estimation with the proposed methods.

The last simulation model considered for the estimation is a bivariate extension of the univariate Ornstein-Uhlenbeck process (BI-OU),

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dB_t, \tag{4.5}$$

where $X_t = (X_{1t}, X_{2t})$, $\kappa = \begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix}$, $\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$ and $\sigma = \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}$. Under the condition that the eigenvalues of the matrix κ have positive real parts, the process is stationary with transition distribution being a bivariate normal $N(m(\delta, X_t), \Omega(\delta))$ where $m(\delta, X_t) = \alpha + exp(-\kappa\delta)(X_t - \alpha)$, $\Omega(\delta) = \Sigma - exp(-\kappa\delta)\Sigma exp(-\kappa^T\delta)$ and

$$\Sigma = \frac{1}{2tr(\kappa)Det(\kappa)}\{Det(\kappa)\sigma\sigma^T + \{\kappa - tr(\kappa)\}\sigma\sigma^T\{\kappa - tr(\kappa)\}^T\}.$$

The CCF of the process is known to be $\psi_t(u_1, u_2; \theta) = exp\{iu^T m(\delta, X_t) - u^T\Omega(\delta)u/2\}$ for $u = (u_1, u_2)^T$.

We then carried out simulations to evaluate the ability of the proposed tests in detecting model deviations. When we chose the simulation models, we had in mind two issues in finance that have drawn considerable research attention recently. The first issue is whether the process is subject to jumps, and the second is whether we could differentiate two processes with different jump rates. Our simulation study formulated two settings of hypotheses to address these two issues. In the first setting, we tested

$$H_0 : \text{The process is the VSK model.}$$

In the second setting, we tested

$$H_0 : \text{The process is the jump diffusion model VSK-MJ.}$$

For computing the powers, in the first setting we used the data simulated from H_1 : the jump diffusion model VSK-MJ to test the null model which does not have jumps; in the second setting, we used the data simulated from H_1 : the Inverse Gaussian OU model which has infinite-activity jumps to test the null hypothesis that prescribes a finite-activity jump process.

For each model, we simulated 500 sample paths which were observed at monthly observations ($\delta = 1/12$) for $n = 125, 250, 500$ respectively. The choices of parameter values were motivated by Chen, Gao and Tang (2008) and Ait-Sahalia, Fan and Peng (2009).

In parameter estimation, we discovered that for both real and imaginary parts of the CCF, their non-parametric smoothing estimators are wave-like functions and roughly diminish to zero at the same points, which creates a region denoted as S_t (here the subscript t indicates that the region depends on X_t). In practice, we searched on a couple of grid points in the data range of X_t and picked the union of S_t as the support region S for the frequency domain of $\psi_t(u;\theta)$ in the estimation. We then chose the uniform density as the weight function π over the support region.

In model testing, similar effort was initially made to obtain the support region of the non-parametric CCF estimate, denoted as S_{NP}, and the support region of the theoretical CCF under H_0, denoted as S_{H_0}. Here the theoretical CCF under H_0 used $\hat{\theta}_n$ from our EL method. Then the support region of the frequency domain in testing was taken as the union of S_{NP} and S_{H_0}. We chose the uniform density as the weight function over this support region for testing. There is little contribution to the integrated empirical likelihood ratio $\ell_{nh}(\hat{\theta}_n)$ from outside the support region. The biweight Kernel $K(u) = 15/16(1-u^2)^2 I(|u| \le 1)$ was used for smoothing in testing. The bandwidth selection is described in Section 3. The bandwidth sets were specified in Tables 3 and 4 for the two test settings. It is observed that the values of the bandwidths were quite small, which was due to the rapid oscillation of the CCF curves which favored smaller bandwidth in the curve fitting.

We chose $w(u,r;X_t) = e^{ir^T X_t}$ throughout our simulation study as it is the optimal instrument suggested in Carrasco, Chernov, Florens and Ghysels (2007). Some numerical exploration (not reported) indicated the choice of the function $w(\cdot)$ is not crucial in the context of the paper. For testing, we picked the unit instrument to reduce computing burden.

Table 1 reports the empirical averages of the parameter estimates and their standard errors as well as the true parameter values used for simulation. When the sample size increases, standard errors of the proposed all estimates decrease, indicating the consistency of the estimators. We observe from Table 1 (a)-(b) for the VSK and CIR models where the MLEs are available, the proposed EL estimates are quite close to the MLEs. Although the EL estimates tend to have larger standard errors than the MLEs, we do note that under the VSK model in Table 1 (a), the bias of EL estimates for the mean reverting parameter κ are smaller than the corresponding MLEs for all $n = 125$, $n = 250$ and $n = 500$. For the jump diffusion model VSK-MJ (Table 1 (c)), we see the EL estimates are consistently more efficient than the approximate MLEs in the estimation of κ and the Poisson intensity λ. For the Inverse Gaussian OU model which does not have the MLE to compare with, the proposed estimates as reported in Table 1 (d) are close to the true values and the standard errors converge as the sample size increases.

Table 2 reports the estimates for the bivariate OU process and shows that the EL estimates are close to the corresponding MLEs, providing the further evidence of the effectiveness of our EL estimator for multivariate process estimation. We also found that the EL estimates for the long run mean α_1 and the volatility σ_{11} of the first process have smaller biases and standard errors than the MLEs for all $n = 125$, $n = 250$ and $n = 500$.

Tables 3 and 4 report the empirical size and power of the proposed test based on $B = 250$ bootstrap resampled paths for each simulation. They contain the sizes and powers for the overall test that is based on the five bandwidth set, and for the tests that only use one bandwidth. We observe that the tests gave satisfactory sizes under both testing settings. In the first test where we used the data from the jump diffusion model VSK-MJ to test the continuous diffusion model VSK, the powers range from 65% to 95% across the different sample sizes and bandwidths. In the second test where we used data simulated from the infinity-activity jump process (the Inverse Gaussian OU) to test the finite-activity jump process (the jump diffusion VSK-MJ), the powers range from 71% to 90% across the different sample sizes and bandwidth choices.

5 A Case Study

In this section, we examine empirically the capability of our testing procedure in detecting jumps using the secondary market quotes of the 3-month Treasury Bill (T-bill) between January 1, 1965 and February 2, 1999. This bill was sampled at monthly frequency, and in total we had 410 observations. The mean of these bills is 0.065, the volatility is 0.026, the mean of the differences is very close to zero (1.5×10^{-5}) and the standard deviation of the differences is 0.005. The sample period contains some large movements that turn out to coincide with arrivals of macroeconomic news (Johannes (2004)). The goal of this empirical study was to test whether the underlying process is subject to jumps or not.

The proposed parameter estimates under each of the four univariate models considered in the simulation study are reported in Table 5. For comparison, the MLEs or the approximate MLEs are also reported except for the Inverse Gaussian OU model. For the univariate diffusion models VSK and CIR, and the jump diffusion model VSK-MJ, the proposed parameter estimates based on CCF are very similar to the MLEs or the approximate MLEs. The EL estimates of the long-run mean α are 0.059 for VSK and 0.064 for CIR, both of which are close to the summary statistic of mean rates (0.065). In VSK, the average volatility of 3-month T-bill monthly return (difference) is estimated to be $\sigma\sqrt{\delta} = 0.018\sqrt{1/12} = 0.005$, which is also close to the summary statistic of volatility for the change (0.005). However the conditional volatility of monthly change in CIR model is $\sigma\sqrt{\delta X_t}$, and X_t has a long-run average 0.064 which is less than 1. Therefore, the process needs to have higher σ (0.057) to bring up the average volatility of monthly change to the same level reflected by the real data. In the jump diffusion model VSK-MJ, our estimate of λ suggests on average about 2 jumps per year. Relative to VSK and CIR models, the estimate for parameter σ in the jump diffusion VSK-MJ model is much smaller (0.008), indicating that allowing jumps in the process helps capturing large movements in the interest rate, and as a result the continuous part of the process does not have to be as volatile as the one in VSK or CIR models.

We then applied the proposed test for the validity of each of the four models. The bandwidth prescribed by the CV was 0.01. By exploring the kernel estimators of the CCF, a reasonable range for h was from 0.01 to 0.018, that offered smoothness from slightly under-smoothing to slightly over-smoothing. The bandwidth range used in our empirical study consisted of five equally spaced bandwidths ranging from 0.01 to 0.018. Table 6 reports p-values of single bandwidth and the overall tests for the four models. There is no empirical support for VSK model. CIR model performs a little bit better as the distances between the test statistics and the critical values decrease, but the model is still rejected at significance level of 0.05 in the overall test and almost all the single bandwidth tests. We can not reject the jump diffusion model VSK-MJ in the overall test and the single bandwidth tests except the one with the smallest bandwidth (p-value = 0.046) . This constitutes a strong evidence for the presence of jumps and implies that adding (finite-activity) jumps does help capturing the underlying dynamics of the interest rates. By allowing the infinite-activity jumps in the models, the p-values of the tests for the Inverse Gaussian OU model are very supportive even for the small bandwidths, suggesting that the infinite-activity jump model might potentially model the dynamics of the 3-month T-bill rates better. A possible reason for it is that the jump diffusion model VSK-MJ can only generate small continuous movements from Brownian motion and big spikes from the compound Poisson component, but it could miss the movements that are between (i.e. the movements with median sizes). However, the Inverse Gaussian OU process is more flexible since it can generate small, median, and big movements with infinite arrival rates, therefore it could fill in a gap in the VSK-MJ model by capturing movements that are too large for Brownian motion to model but too small for the compound Poisson process to capture.

Appendix

The following conditions are required in our analysis.

C1: The stochastic processes given in (1.1) and (3.1) admit unique weak solution respectively, which are $\alpha-$mixing with mixing coefficient $\alpha(t) = Ce^{-\lambda t}$ where $\alpha(t) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \Omega_1^s, B \in \Omega_{s+t}^\infty\}$ for all $s, t \geq 1$, where C is a finite positive constant and Ω_i^j denotes the σ-field generated by $\{X_t : i \leq t \leq j\}$.

C2: (Smoothness) $\psi_t(\tau;\theta) =: \psi(\tau;\theta, X_t)$ and $E\{\epsilon_t(\tau;\theta)\}$ are third continuous differentiable with respect to θ within a neighborhood of θ_0 which is defined in C3. $\pi(\cdot)$ is a bounded probability density supported on a compact set $S \subset R^d$; and the diffusion function $\sigma(x)$ is positive definite.

C3: The parameter space Θ is an open subset of R^p, and the true parameter θ_0 is the unique root of $E\{\epsilon_t(\tau;\theta)\} = 0$ for all $\tau \in S$; and for any $\theta_1 \neq \theta_2$, $P\left\{\psi_t(\cdot;\theta_1) \neq \psi_t(\cdot;\theta_2, X_t)\right\} > 0$.

C4: (Invertibility) The Hermitian matrix $Var\{\tilde{\epsilon}_t(\tau;\theta_0)\}$ is positive definite almost everywhere for $\tau \in R^{2d}$ with respect to the Lebesgue measure in R^{2d}; $\Gamma(\theta_0)$ defined in (2.9) is invertible.

C5: The kernel $K(\cdot)$ is a r-th order symmetric kernel supported on $[-1,1]^d$ and has bounded second derivative. We assume $d < 4$ and the smoothing bandwidth $h = O\{n^{-1/(d+2r)}\}$. The bandwidth set $\{h_1, \cdots, h_k\}$ satisfies $h_i = c_i h$ for constants c_i such that $c_1 < c_2 < ... < c_k$ where k is an integer not depending on n.

C6: $\{\Delta_n(u;X_t)\}$ is a sequence of complex functions continuous at $u = 0$ and $\Delta_n(0;X_t) \equiv 0$, $\sup_n |\Delta_n(u;X_t)| \leq M_1$ almost surely and the Lebesgue measure of $\{u|\Delta_n(u,x) \neq 0\}$ is positive for all x in the support of the marginal density f, and $c_n = n^{-1/2}h^{-d/4}$ which is the order of the difference between H_0 and H_1.

We need C1 as the basic condition for the stochastic processes involved. Ait-Sahalia (1996) and Genon-Catalot, Jeantheau and Laredo (2000) provides conditions on the underlying processes such that the Assumption C1 held. In particular, Ait-Sahalia (1996) provides conditions so that the observed sequences are β-mixing, which is automatically α-mixing. We require the rate of decay is exponentially fast to simplify the technical arguments. C2 consists of smoothness conditions regarding the CCFs and C3 is for identification of parameters. C4 ensures the covariance matrix is invertible which is easier to be justified for our low dimensional formulation of estimation and testing approaches. C5 on the kernel and bandwidth are standard in non-parametric curve estimation. The assumption of $d < 4$ is to make the bias in the kernel estimation a smaller order of $h^{d/2}$ so that the bias is stochastically negligible relative to $\ell_{nh}(\theta_0)$. The kernel method will encounter the curse of dimensionality when $d \geq 4$. Also, the commonly used processes in finance and other stochastic modeling tend to have dimension less than 4. The bandwidth selected by either cross validation or the plug-in method satisfies the order specified in C5. The first part of C6 regarding $\Delta_n(u; X_t)$ is to qualify $\psi_t(u;\theta)$ under H_1 as a bona fide characteristic function, whereas the part that requires positive measure on the set $\{u|\Delta_n(u,x) \neq 0\}$ is to make H_1 a genuine sequence of alternative hypotheses.

Proof of Lemma 1. By combining results in Kitamura (1997) and Chen, Härdle and Li (2003) for the empirical likelihood of $\alpha-$mixing processes, we can show that

$$\lambda(\tau;\theta) = A_n^{-1}(\tau;\theta)\{\frac{1}{n}\sum_{t=1}^{n}\vec{\epsilon}_t(\tau;\theta)\} + o(n^{-1/3}) = O(n^{-1/3}) \tag{A.1}$$

almost surely and uniformly in $||\theta - \theta_0|| \leq n^{-1/3}$ and $\tau^T \in S$. Denote $\theta = \theta_0 + un^{-1/3}$. It follows from (A.1) and Taylor expansion that, uniformly in $||u|| = 1$,

$$\begin{array}{rl} & \ell_n(\theta) \\ = & \int\{2\sum_{t=1}^n \lambda^T(\tau;\theta)\vec{\epsilon}_t(\tau;\theta) - \sum_{t=1}^n\{\lambda^T(\tau;\theta)\vec{\epsilon}_t(\tau;\theta)\}^2\}\pi(\tau)\,d\tau + o(n^{1/3}) \\ = & \int n\{\frac{1}{n}\sum_{t=1}^n \vec{\epsilon}_t^T(\tau;\theta_0) + \frac{1}{n}\sum_{t=1}^n \frac{\partial\vec{\epsilon}_t^T(\tau;\theta_0)}{\partial\theta}un^{-1/3}\}A_n^{-1}(\tau;\theta) \\ & \quad\times\{\frac{1}{n}\sum_{t=1}^n \vec{\epsilon}_t^T(\tau;\theta_0) + \frac{1}{n}\sum_{t=1}^n \frac{\partial\vec{\epsilon}_t^T(\tau;\theta_0)}{\partial\theta}un^{-1/3}\}\pi(\tau)\,d\tau + o(n^{1/3}) \\ = & \int n\{E(\frac{\partial\vec{\epsilon}_1^T(\tau;\theta_0)}{\partial\theta})un^{-1/3}(1+o(1))\}A^{-1}(\tau,\tau;\theta_0,\theta_0) \\ & \times\{E(\frac{\partial\vec{\epsilon}_1(\tau;\theta_0)}{\partial\theta})un^{-1/3}(1+o(1))\}\pi(\tau)\,d\tau + o(n^{1/3}) \\ \geq & \frac{1}{2}cn^{1/3} \end{array} \tag{A.2}$$

almost surely, where $c > 0$ is the smallest eigenvalue of

$$\sup_{\tau\in S} E(\frac{\partial\vec{\epsilon}_1^T(\tau;\theta_0)}{\partial\theta})A^{-1}(\tau,\tau;\theta_0,\theta_0)E(\frac{\partial\vec{\epsilon}_1(\tau;\theta_0)}{\partial\theta}).$$

Similarly,

$$\begin{array}{rcl} \ell_n(\theta_0) & = & \int\{\sum_{t=1}^n \vec{\epsilon}_t^T(\tau;\theta_0)\}A^{-1}(\tau,\tau;\theta_0,\theta_0)\{\frac{1}{n}\sum_{t=1}^n \vec{\epsilon}_t(\tau;\theta_0)\}\pi(\tau)\,d\tau + o(1) \\ & = & o(n^{1/3}) \end{array} \tag{A.3}$$

almost surely. This together with (A.2) implies that $\ell_n(\theta)$ has a minimum value in the interior of the ball $||\theta - \theta_0|| \leq n^{-1/3}$ and this value satisfies $\frac{\partial}{\partial\theta}\ell_n(\theta) = 0$, i.e., the second equation in (2.7) by noting (2.4). The first equation follows directly from (2.4).

Proof of Theorem 1. It follows from limit theorems for martingale difference that

$$\left\{\begin{array}{l} \frac{\partial}{\partial\theta}Q_{1n}(\tau;\theta_0,0) = \frac{1}{n}\sum_{t=1}^n \frac{\partial}{\partial\theta}\vec{\epsilon}_t(\tau;\theta_0) \xrightarrow{p} M_0E\{\frac{\partial}{\partial\theta}\tilde{\epsilon}_1(\tau;\theta_0)\} \\ \frac{\partial}{\partial\lambda^T}Q_{1n}(\tau;\theta_0,0) = -\frac{1}{n}\sum_{t=1}^n \vec{\epsilon}_t(\tau;\theta_0)\vec{\epsilon}_t^T(\tau;\theta_0) \xrightarrow{p} -M_0A(\tau,\tau;\theta_0,\theta_0)M_0^\star \\ \frac{\partial}{\partial\theta}Q_{2n}(\tau;\theta_0,0) = 0 \\ \frac{\partial}{\partial\lambda^T}Q_{2n}(\tau;\theta_0,0) = \frac{1}{n}\sum_{t=1}^n \frac{\partial}{\partial\theta}\vec{\epsilon}_t^T(\tau;\theta_0) \xrightarrow{p} E\{\frac{\partial}{\partial\theta}\tilde{\epsilon}_1^\star(\tau;\theta_0)\}M_0^\star \end{array}\right. \tag{A.4}$$

uniformly in $\tau^T \in S$. Put $\delta_n = ||\hat{\theta}_n - \theta_0|| + \sup_{\tau^T\in S}||\lambda(\tau;\hat{\theta}_n)||$. Then it follows from Taylor expansion that

$$\begin{array}{rl} 0 & = Q_{1n}(\tau;\hat{\theta}_n,\lambda(\tau;\hat{\theta}_n)) \\ & = Q_{1n}(\tau;\theta_0,0) + \frac{\partial Q_{1n}(\tau;\theta_0,0)}{\partial\theta}(\hat{\theta}_n - \theta_0) + \frac{\partial Q_{1n}(\tau;\theta_0,0)}{\partial\lambda^T}\lambda(\tau;\hat{\theta}_n) + o_p(\delta_n) \end{array} \tag{A.5}$$

uniformly in $\tau^T \in S$, and

$$\begin{array}{rcl} 0 & = & \int Q_{2n}(\tau;\hat{\theta}_n,\lambda(\tau;\hat{\theta}_n))\pi(\tau)\,d\tau \\ & = & \int\{Q_{2n}(\tau;\theta_0,0) + \frac{\partial Q_{2n}(\tau;\theta_0,0)}{\partial\theta}(\hat{\theta}_n - \theta_0) + \frac{\partial Q_{2n}(\tau;\theta_0,0)}{\partial\lambda^T}\lambda(\tau;\hat{\theta}_n)\}\pi(\tau)\,d\tau \\ & & +o_p(\delta_n). \end{array} \tag{A.6}$$

By (A.4) - (A.6), we have

$$\begin{array}{rl} & \hat{\theta}_n - \theta_0 \\ = & -\Gamma^{-1}(\theta_0) \int E\{\frac{\partial}{\partial\theta}\tilde{\epsilon}_1^\star(\tau;\theta_0)\} A^{-1}(\tau;\theta_0,\theta_0) M_0^{-1} \frac{1}{n}\sum_{t=1}^n \vec{\epsilon}_t(\tau;\theta_0)\pi(\tau)\, d\tau + o_p(\delta_n). \end{array} \tag{A.7}$$

Hence the theorem follows from (A.7) and the central limit theorem for Martingale difference.

Proof of Theorem 2. Define $V(\tau_1,\tau_2,x;\theta_0,\theta) = E\{\tilde{\epsilon}(\tau_1,X_t;\theta)\tilde{\epsilon}^\star(\tau_2,X_t;\theta)|X_t = x\}$ and write $V(\tau,x;\theta_0,\theta) = V(\tau,\tau,x;\theta_0,\theta)$. Since $\hat{\theta}_n$ is $\sqrt{n}$-consistent to θ_0, we have

$$\begin{array}{rcl} \ell_{nh}(\hat{\theta}_n) & = & \ell_{nh,1}(\theta_0) + nh^d R^{-1}(K)\{(\hat{\theta}-\theta_0)^T S_{n,h}(\theta_0) + S_{n,h}^\star(\theta_0)(\hat{\theta}_n - \theta_0) \\ & & +(\hat{\theta}_n-\theta_0)^T\Gamma_{n,h}(\theta_0)(\hat{\theta}_n-\theta_0)\} + O_p\{(nh^d)^{-1/2}\log^3(n) \\ & & +h^2\log^2(n)\} \end{array} \tag{A.8}$$

where

$$\begin{array}{rcl} \ell_{nh,1}(\theta_0) & = & nh^d R^{-1}(K) \int\int \tilde{\epsilon}_{n,h}^\star(\tau,X_t;\theta_0) V^{-1}(\tau,x;\theta_0,\theta_0) \\ & & \times\tilde{\epsilon}_{n,h}(\tau,x;\theta_0)\pi_1(\tau) f^{-1}(x)\pi_2(x) d\tau dx, \end{array} \tag{A.9}$$

$$\begin{array}{rcl} S_{n,h}(\theta_0) & = & \int\int \frac{\partial\tilde{\epsilon}_{n,h}^\star(\tau,x;\theta_0)}{\partial\theta} V^{-1}(\tau,x;\theta_0,\theta_0)\tilde{\epsilon}_{n,h}(\tau,x;\theta_0) \\ & & \times\pi_1(\tau)\pi_2(x) f^{-1}(x) d\tau dx, \end{array} \tag{A.10}$$

$$\begin{array}{rcl} \Gamma_{nh}(\theta_0) & = & \int\int \frac{\partial\tilde{\epsilon}_{n,h}^\star(\tau,x;\theta_0)}{\partial\theta} V^{-1}(\tau,x;\theta_0,\theta_0)\frac{\partial\tilde{\epsilon}_{n,h}(\tau,x;\theta_0)}{\partial\theta} \\ & & \times\pi_1(\tau)\pi_2(x) f^{-1}(x) d\tau dx. \end{array} \tag{A.11}$$

As $S_{n,h}(\theta_0) = O_p(n^{-1/2})$,

$$\ell_{nh}(\hat{\theta}_n) = \ell_{nh,1}(\theta_0) + O_p\{(nh^d)^{-1/2}\log^3(n) + h^2\log^2(n) + h^d\}. \tag{A.12}$$

Note that

$$\begin{array}{rl} & \ell_{nh,1}(\theta_0) \\ = & nh^d R^{-1}(K)\int\int n^{-1}\sum_{t_1=1}^n K_h(x-X_{t_1})\{\tilde{\epsilon}^\star(\tau,X_{t_1}) + c_n\tilde{\eta}_n^\star(\tau,X_{t_1})\} \\ & \times V^{-1}(\tau,x;\theta_0,\theta_0) n^{-1}\sum_{t_2=1}^n K_h(x-X_{t_2})\{\tilde{\epsilon}(\tau,X_{t_2}) + c_n\tilde{\eta}_n(\tau,X_{t_2})\} \\ & \times\pi_1(\tau)\pi_2(x) f^{-1}(x) d\tau dx + o_p(h^{d/2}) \\ = & R^{-1}(K)\left(H_{n1}+H_{n2}+H_{n3}+H_{n4}\right) + o_p(h^{d/2}), \end{array} \tag{A.13}$$

where, with the choice of $c_n = n^{-1/2}h^{-d/4}$,

$$\begin{array}{rcl} H_{n1} & = & n^{-1}h^d \sum_{t_1\neq t_2}\int\int K_h(x-X_{t_1})K_h(x-X_{t_2})\tilde{\epsilon}^\star(\tau,X_{t_1})V^{-1}(\tau,x) \\ & & \times\tilde{\epsilon}(\tau,X_{t_2})\pi_1(\tau)\pi_2(x) f^{-1}(x) d\tau dx, \\ H_{n2} & = & n^{-1}h^d\sum_{t=1}^n\int\int K_h^2(x-X_t)\tilde{\epsilon}^\star(\tau,X_t)V^{-1}(\tau,x)\tilde{\epsilon}(\tau,X_t) \\ & & \times\pi_1(\tau)\pi_2(x) f^{-1}(x) d\tau dx, \\ H_{n3} & = & 2n^{1/2}h^{3d/4}\int\int\tilde{\eta}_n^\star(\tau,x)V^{-1}(\tau,x) n^{-1}\sum_{t=1}^n K_h(x-X_t)\tilde{\epsilon}(\tau,X_t) \\ & & \times\pi_1(\tau)\pi_2(x) f^{-1}(x) d\tau dx, \\ H_{n4} & = & h^{d/2}\int\int\tilde{\eta}_n^\star(\tau,x)V^{-1}(\tau,x)\tilde{\eta}_n(\tau,x)\pi_1(\tau)\pi_2(x) f^{-1}(x) d\tau dx. \end{array} \tag{A.14}$$

We note that $H_{n2} = 2R(K) + o_p(h^d)$ and and the integral in H_{n3} is $O_p(n^{-1/2})$. Hence, $H_{n3} = O_p(n^{3d/4}) = o_p(h^{d/2})$.

Now consider H_{n1}. Clearly, $E(H_{n1}) = 0$ and the double summation in H_{n1} constitutes a generalized U-statistic of order two with the kernel

$$\begin{aligned}\xi_{t_1,t_2} &= \int\int K_h(x - X_{t_1})K_h(x - X_{t_2})\tilde{\epsilon}^\star(\tau, X_{t_1})V^{-1}(\tau, x; \theta_0, \theta_0)\tilde{\epsilon}(\tau, X_{t_2}) \\ &\quad \times\pi_1(\tau)\pi_2(x)f^{-1}(x)d\tau dx.\end{aligned} \tag{A.15}$$

The U-statistic is degenerate due to $\{\tilde{\epsilon}(\tau, X_{t_2})\}$ being martingale differences.

Let $\sigma_n^2 = \sum_{1\le t_1\ne t_2\le n} \sigma_{t_1,t_2}^2$ where $\sigma_{t_1,t_2}^2 = Var(\xi_{t_1,t_2})$. Then, apply the central limit theorem for generalized U-statistics for α-mixing sequences (Gao and King, 2005), we have

$$\sigma_n^{-1} \sum_{t_1\ne t_2} \xi_{t_1,t_2} \xrightarrow{d} N(0, 1). \tag{A.16}$$

Furthermore, it can be shown, for instance by following the route of Chen, Gao and Tang (2008) that $\sigma_n^2 = 2n^2\sigma_{n0}^2\{1 + o(1)\}$ where $\sigma_{n0}^2 = E_{t_1}E_{t_2}(\xi_{t_1,t_2}^2)$. Here E_{t_i} denote marginal expectation with respect to (X_{t_i}, X_{t_i+1}).

It can be shown that

$$\begin{aligned}\sigma_{n0}^2 &= \int\int\int\int E_{t_1}E_{t_2}\{K_h(x_1 - X_{t_1})K_h(x_1 - X_{t_2})K_h(x_2 - X_{t_1}) \\ &\quad \times K_h(x_2 - X_{t_2}) \textstyle\sum_{l_1,k_1,l_2,k_2}^2 \epsilon_{l_1}(\tau_1, X_{t_1})\epsilon_{k_1}(\tau_1, X_{t_2})\epsilon_{l_2}(\tau_2, X_{t_1}) \\ &\quad \times\epsilon_{k_2}(\tau_2, X_{t_2})\nu^{l_1,k_1}(\tau_1, x_1)\nu^{l_2,k_2}(\tau_2, x_2)\} \\ &\quad \times\pi_1(\tau_1)\pi_1(\tau_2)f^{-1}(x_1)f^{-1}(x_2)\pi_2(x_1)\pi_2(x_2)d\tau_1 d\tau_2 dx_1 dx_2 \\ &= \int\int\int\int E_{t_1}E_{t_2}\{K_h(x_1 - X_{t_1})K_h(x_1 - X_{t_2})K_h(x_2 - X_{t_1}) \\ &\quad \times K_h(x_2 - X_{t_2}) \textstyle\sum_{l_1,k_1,l_2,k_2}^2 V_{l_1l_2}(-\tau_1, \tau_2, X_{t_1})V_{k_1k_2}(\tau_1, -\tau_2, X_{t_2}) \\ &\quad \times\nu^{l_1,k_1}(\tau_1, x_1)\nu^{l_2,k_2}(\tau_2, x_2)\}\pi_1(\tau_1)\pi_1(\tau_2)f^{-1}(x_1)f^{-1}(x_2) \\ &\quad \times\pi_2(x_1)\pi_2(x_2)d\tau_1 d\tau_2 dx_1 dx_2 \\ &= h^{-d}\gamma^2(K, V, \pi_1, \pi_2)\{1 + O(h^2)\},\end{aligned} \tag{A.17}$$

where $\gamma^2(K, V, \pi_1, \pi_2)$ is defined in (3.4). From (A.16) and (A.17), we have

$$h^{-d/2}H_{n1} \xrightarrow{d} N(0, 2\gamma^2(K, V, \pi_1, \pi_2)) \tag{A.18}$$

This together with the results on H_{n2} and H_{n3} leads to

$$h^{-d/2}(\ell_{nh}(\hat{\theta}) - 2 - \mu_n) \xrightarrow{d} N(0, 2R^{-2}(K)\gamma^2(K, V, \pi_1, \pi_2)) \tag{A.19}$$

where $\mu_n = H_{n4}$. This completes the proof of Theorem 2.

Proof of Theorem 3. The proof can be made by applying the Cramér-Wold device and the same technique in the proof of Theorem 2 followed by the mapping theorem.

Acknowledgements:

The authors thank two reviewers and an associate editor for helpful comments, and acknowledge support from National Science Foundation grants DMS-0604563, DMS-0518904, SES-0631608 and DMS-1005336. The first author was partially supported through Center for Statistical Science at Peking University.

References

[1] Aït-Sahalia, Y. (1996). Nonparametric pricing of interest rate derivative securities. *Econometrica* **64**, 527–560.

[2] Aït-Sahalia, Y. (2002), Maximum-likelihood estimation of discretely-sampled diffusions: a closed-form approximation approach. *Econometrica* **70**, 223 – 262.

[3] Aït-Sahalia, Y. (2008), Closed-form likelihood expansions for multivariate diffusions. *Ann. Statist.* **36**, 906 – 937.

[4] Aït-Sahalia, Y., Fan, J. and Peng, H. (2009), Nonparametric transition-based tests for jump-diffusions. J. Amer. Statist. Asso. **104**, 1102–1116.

[5] Barndorff-Nielsen, O. E. and Shephard, N. (2001), Non-Gaussian Orstein-Uhlenbeck-based models and some of their uses in financial econometrics. *Journal of the Royal Statistical Society, Series B* **63**, 167 – 241.

[6] Carrasco, M., Chernov, M., Florens, J-P. and Ghysels, E. (2007), Efficient estimation of jump diffusion and general dynamic models with a continuum of moment conditions. *Journal of Econometrics* **140**, 529 – 573.

[7] Chacko, G. and Viceira, L. M. (2003), Spectral GMM estimation of continuous-time processes, *Journal of Econometrics* **116, no. 1-2**, 259 – 292.

[8] Chen, B. and Hong, Y. (2010), Characteristic function-based testing for multifactor continuous-Time markov models via nonparametric regression. Econometric Theory **26**, 1115–1179.

[9] Chen, S. X., Gao, J. and Tang, C.Y. (2008), A test for model specification of diffusion models. Ann. of Statist. **36**, 167 – 198.

[10] Chen, S. X., Härdle, W. and Li, M. (2003), An empirical likelihood goodness–of–fit test for time series. Journal of the Royal Statistical Society, Series B 65, 663 – 678.

[11] Chen, S. X. and I. Van Keilegom (2009), A review on empirical likelihood for regressions (with discussions), Test, **3**, 415-447 .

[12] Chen, S. X., Peng, L. and Yu, C. L. (2011), Supplement to "Parameter Estimation and Model Testing for Markov Processes via Conditional Characteristic Functions". DOI: XXXX.

[13] Cox, J. C., Ingersoll, J.E., and Ross, S.A. (1985), A theory of term structure of interest Rates. Econometrica **53**, 385 – 407.

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

[15] Fan, J. (2005), A selective overview of nonparametric methods in financial econometrics. Statistical Science **20**, 317 – 357.

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

[17] Feuerverger, A. and McDunnough, P. (1981), On some Fourier methods for inference. J. Amer. Statist. Asso. **76**, 379 – 387.

[18] Feuerverger, A. and Mureika, R. A. (1977), The empirical characteristic function and its applications. Ann. of Statist. **5**, 88 – 97.

[19] Genon-Catalot, V., Jeantheau, T. and Laredo, C. (2000). Stochastic volatility models as hidden Markov models and statistical applications. Bernoulli **6**, 1051C1079.

[20] Gao, J. and King, M. (2005), Estimation and model specification testing in nonparametric and semiparametric regression models. Unpublished paper available at www.maths.uwa.edu.au/~jiti/jems.pdf.

[21] Jiang, G. J. and Knight, J. L. (2002), Estimation of continuous-time processes via the empirical characteristic function. Journal of Business and Economic Statistics **20**, 198 – 212.

[22] Johannes, M. (2004), The statistical and economic role of jumps in continuous-time interest rate models. Journal of Finance **1**, 227 – 260.

[23] Kitamura, Y. (1997), Empirical likelihood methods with weakly dependent processes. Ann. of Statist. **25**, 2084 – 2102.

[24] Kitamura, Y., Tripathi, G. and Ahn, H. (2004), Empirical likelihood-based inference in conditional moment restriction models. Econometrica **72**, 16671714.

[25] Merton,R. (1976), Option pricing when the underlying stock returns are discontinuous. Journal of Financial Economics **3**, 125 – 144.

[26] Øksendal, B. (2000). Stochastic Differential Equations: An Introduction with Applications, Fifth Edition, Berlin: Springer.

[27] Owen, A. (1988), Empirical likelihood ratio confidence ratios for a single functional. *Biometrika* **75**, 237 – 249.

[28] Singleton, K. J. (2001), Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics* **102**, 111 – 141.

[29] Sørensen, M. (1991), Likelihood methods for diffusions with jumps. In *Statistical Inference in Stochastic Processes,* Ed. by Prabhu, N. U. and Basawa, I. V., 67 - 105.

[30] Strook, D. and Varadhan, S. (1979), *Multivariate Diffusion Processes*, Springer.

[31] Sundaresan, S. M. (2000), Continuous time finance: A review and assessment. Journal of Finance **55**, 1569 – 1622.

[32] Vasicek, O. (1977), An equilibrium characterization of the term structure. Journal of Financial Economics **5**, 177 – 188.

Table 1: Empirical averages and their standard errors (in parentheses) of the maximum (MLE) or approximate maximum (AMLE) likelihood estimates and the proposed empirical likelihood estimates (EL) under the four univariate models.

(a) Vasicek Model

n		$\kappa = 0.858$	$\alpha = 0.089$	$\sigma = 0.047$
125	MLE	1.383(0.603)	0.090(0.015)	0.047(0.003)
	EL	1.305(0.643)	0.090(0.017)	0.046(0.004)
250	MLE	1.118(0.397)	0.090(0.011)	0.047(0.002)
	EL	1.052(0.410)	0.089(0.013)	0.046(0.002)
500	MLE	0.966(0.240)	0.089(0.008)	0.047(0.002)
	EL	0.951(0.273)	0.089(0.009)	0.047(0.002)

(b) CIR Model

n		$\kappa = 0.892$	$\alpha = 0.091$	$\sigma = 0.181$
125	MLE	1.372(0.644)	0.091(0.019)	0.183(0.012)
	EL	1.290(0.719)	0.093(0.023)	0.178(0.014)
250	MLE	1.127(0.374)	0.090(0.013)	0.182(0.008)
	EL	1.089(0.435)	0.091(0.015)	0.179(0.009)
500	MLE	1.000(0.245)	0.091(0.010)	0.182(0.006)
	EL	0.977(0.290)	0.092(0.011)	0.180(0.007)

(c) Jump Diffusion VSK-MJ Model

n		$\kappa = 0.858$	$\alpha = 0.089$	$\sigma = 0.047$	$\lambda = 2.0$	$\eta = 0.067$
125	AMLE	1.056(0.381)	0.093(0.020)	0.046(0.005)	1.770(0.723)	0.060(0.016)
	EL	1.090(0.261)	0.084(0.031)	0.048(0.009)	1.851(0.323)	0.066(0.020)
250	AMLE	0.977(0.226)	0.093(0.013)	0.047(0.003)	1.659(0.466)	0.059(0.010)
	EL	1.043(0.201)	0.090(0.023)	0.048(0.007)	1.825(0.236)	0.068(0.015)
500	AMLE	0.939(0.145)	0.092(0.009)	0.047(0.002)	1.620(0.311)	0.060(0.007)
	EL	1.018(0.115)	0.089(0.018)	0.049(0.005)	1.801(0.163)	0.068(0.012)

(d) Inverse Gaussian OU Model

n		$\lambda = 10.0$	$a = 1.0$	$b = 20.0$
125	EL	10.328(3.665)	1.048(0.106)	20.722(2.146)
250	EL	11.154(1.976)	1.059(0.043)	21.380(0.878)
500	EL	11.489(1.652)	1.031(0.024)	20.846(0.461)

Table 2: Empirical averages and their standard errors (in parentheses) of the maximum (MLE) likelihood estimates and the proposed empirical likelihood estimates (EL) under the Bivariate OU model.

n		$\kappa_{11} = 0.22$	$\kappa_{21} = 0.2$	$\kappa_{22} = 0.5$
125	MLE	0.441(0.197)	0.395(0.270)	0.607(0.176)
	EL	0.381(0.208)	0.525(0.238)	0.594(0.192)
250	MLE	0.353(0.165)	0.307(0.148)	0.563(0.110)
	EL	0.354(0.178)	0.449(0.184)	0.564(0.153)
500	MLE	0.280(0.118)	0.241(0.104)	0.526(0.068)
	EL	0.261(0.168)	0.383(0.154)	0.487(0.112)

n		$\alpha_1 = 0.08$	$\alpha_2 = 0.09$	$\sigma_{11} = 0.09$	$\sigma_{22} = 0.17$
125	MLE	0.145(0.166)	0.099(0.056)	0.167(0.067)	0.080(0.079)
	EL	0.141(0.141)	0.117(0.085)	0.129(0.044)	0.071(0.034)
250	MLE	0.141(0.151)	0.096(0.036)	0.140(0.065)	0.116(0.074)
	EL	0.142(0.129)	0.094(0.073)	0.095(0.033)	0.094(0.028)
500	MLE	0.102(0.120)	0.092(0.023)	0.115(0.051)	0.146(0.055)
	EL	0.099(0.108)	0.104(0.064)	0.077(0.024)	0.105(0.028)

Table 3: H_0: VSK versus H_1: the jump diffusion model VSK-MJ

(a) Size Evaluation (in percentage)

n=125	Bandwidth	0.012	0.017	0.021	0.025	0.030	Overall
	Size	4.6	5.6	5.4	5.8	5.6	4.8
n=250	Bandwidth	0.012	0.015	0.018	0.021	0.024	Overall
	Size	5.6	6.2	6.2	6.0	5.8	5.4
n=500	Bandwidth	0.011	0.013	0.015	0.018	0.020	Overall
	Size	5.0	5.6	5.6	5.4	5.6	5.0

(b) Power Evaluation (in percentage)

n=125	Bandwidth	0.016	0.021	0.026	0.032	0.037	Overall
	Power	72.0	71.6	70.4	69.2	65.8	72.2
n=250	Bandwidth	0.016	0.019	0.022	0.026	0.029	Overall
	Power	82.4	82.4	82.2	82.4	82.2	82.6
n=500	Bandwidth	0.014	0.017	0.019	0.021	0.024	Overall
	Power	95.0	94.8	94.6	94.4	94.2	94.8

Table 4: H_0: the jump diffusion model VSK-MJ versus H_1:the Inverse Gaussian OU model

(a) Size Evaluation (in percentage)

n=125	Bandwidth	0.017	0.022	0.028	0.034	0.040	Overall
	Size	3.4	3.6	4.0	3.6	4.6	4.6
n=250	Bandwidth	0.017	0.021	0.024	0.028	0.032	Overall
	Size	4.6	4.6	4.6	4.6	5.0	4.8
n=500	Bandwidth	0.016	0.019	0.021	0.024	0.026	Overall
	Size	5.0	5.2	5.2	5.0	5.0	5.0

(b) Power Evaluation (in percentage)

n=125	Bandwidth	0.008	0.012	0.017	0.021	0.026	Overall
	Power	71.6	73.8	73.2	71.4	71.2	74.4
n=250	Bandwidth	0.008	0.011	0.014	0.017	0.020	Overall
	Power	84.0	84.2	83.4	81.8	81.4	84.4
n=500	Bandwidth	0.008	0.010	0.012	0.014	0.016	Overall
	Power	90.1	88.9	89.5	85.1	85.4	90.2

Table 5: Empirical Estimation for the 3-month T-bill Data

(a) VSK Model

	κ	α	σ
MLE	0.277	0.065	0.019
	(0.1800)	(0.0117)	(0.0007)
EL	0.274	0.059	0.018
	(0.1956)	(0.0136)	(0.0007)

(b) CIR Model

	κ	α	σ
MLE	0.182	0.066	0.061
	(0.1697)	(0.0179)	(0.0021)
EL	0.182	0.064	0.057
	(0.1934)	(0.0374)	(0.0021)

(c) VSK-MJ Model

	κ	α	σ	λ	η
AMLE	0.071	0.077	0.009	1.863	0.012
	(0.0170)	(0.0129)	(0.0004)	(0.3282)	(0.0015)
EL	0.072	0.076	0.008	1.862	0.013
	(0.0143)	(0.0136)	(0.0008)	(0.1569)	(0.0021)

(d) Inverse Gaussian OU Model

	λ	a	b
EL	0.264	1.139	12.558
	(0.0342)	(0.1364)	(0.8970)

Table 6: P-values for the 3-month T-bill Data

	Bandwidth	0.010	0.012	0.014	0.016	0.018	Overall
VSK	Test Stats	21.971	19.225	16.145	13.267	10.786	14.828
	$l^*_{0.05}$	3.228	3.123	2.845	2.724	2.647	1.462
	P-values	0.0	0.0	0.0	0.0	0.0	0.0
CIR	Test Stats	6.015	4.775	3.755	2.954	2.335	3.546
	$l^*_{0.05}$	2.782	2.739	2.825	2.650	2.448	1.229
	P-values	0.0	0.01	0.02	0.026	0.054	0.0
VSK-MJ	Test Stats	37.204	40.901	45.046	49.878	55.561	25.600
	$l^*_{0.05}$	35.669	43.548	52.247	62.744	74.298	28.751
	P-values	0.046	0.074	0.102	0.126	0.148	0.0880
IG-OU	Test Stats	10.716	9.374	7.962	6.663	5.528	6.870
	$l^*_{0.05}$	40.463	47.665	46.444	42.396	41.750	27.940
	P-values	0.11	0.148	0.124	0.128	0.122	0.162