The Econometrics Journal

Econometrics Journal (2010), volume **13**, pp. 245–270.
doi: 10.1111/j.1368-423X.2010.00316.x

ECF estimation of Markov models where the transition density is unknown

GEORGE J. JIANG[†,‡] AND JOHN L. KNIGHT[§]

[†]*Department of Finance, Eller College of Management, University of Arizona, 1130 E. Helen Street, Tucson, AZ 85721-0108, USA.*
E-mail: gjiang@eller.arizona.edu

[‡]*Yunnan University of Finance and Economics, 273 Longquan Road, Kunming, China, 650221.*

[§]*Department of Economics, University of Western Ontario, 1151 Richmond St., London, ON N6A 5C2, Canada.*
E-mail: jknight@uwo.ca

First version received: September 2006; final version accepted: January 2010

Summary In this paper, we consider the estimation of Markov models where the transition density is unknown. The approach we propose is based on the empirical characteristic function estimation procedure with an approximate optimal weight function. The approximate optimal weight function is obtained through an Edgeworth/Gram–Charlier expansion of the logarithmic transition density of the Markov process. We derive the estimating equations and demonstrate that they are similar to the approximate maximum likelihood estimation (AMLE). However, in contrast to the conventional AMLE our approach ensures the consistency of the estimator even with the approximate likelihood function. We illustrate our approach with examples of various Markov processes. Monte Carlo simulations are performed to investigate the finite sample properties of the proposed estimator in comparison with other methods.

Keywords: *Approximate MLE*, *Edgeworth/Gram–Charlier expansion*, *Empirical characteristic function (ECF)*, *Markov process*.

1. INTRODUCTION

The estimation of Markov models can proceed straightforwardly via the maximum likelihood estimation (MLE) method when the transition density is known in closed form. However, when the transition density is unknown alternative estimators to MLE need to be considered. Various estimation methods have been proposed and applied in the literature, e.g. the quasi-maximum likelihood (QML) method by Bollerslev and Wooldridge (1992), Fisher and Gilles (1996) among others, the maximum likelihood method based on closed-form Hermite expansion of the likelihood function by Aït-Sahalia (2002, 2003), the generalized method of moments (GMM) by Hansen and Scheinkman (1996) and Liu (1997), as well as various simulation-based methods such as the simulated moments estimator (SME) by Duffie and Singleton (1993), the indirect inference approach by Gouriéroux et al. (1993), the efficient method of moments (EMM) by

Gallant and Tauchen (1996) and Gallant and Long (1997) with applications by Chernov and Ghysels (2000) and Andersen et al. (2002) to continuous-time asset return models, and the Bayesian MCMC method by Jones (1998), etc.

In recent literature, a number of estimation methods have also been developed in the frequency domain. These methods exploit the analytical conditional characteristic function (CCF) of the state variables. It is noted that, for many Markov processes, while a closed form of the transition density is unavailable, the associated CCF of the state variables can often be derived analytically. For instance, in the continuous time finance literature, models specified within the affine framework often have closed-form CCF. This observation opens the door to alternative estimation methods using the CCF. In this regard we have the estimators developed in Singleton (2001), Jiang and Knight (2002) and Chacko and Viceira (2003) for continuous time diffusion and jump-diffusion models. The basic idea is to minimize the integrated distance between the empirical characteristic function (ECF) or joint ECF and their theoretical counterparts. Singleton (2001) also proposes to use the likelihood function obtained by Fourier inversion of the CCF. In a recent paper, Carrasco et al. (2007) extend the ECF approach by using a continuum of moment conditions for the estimation of general dynamic models.

The ECF estimation method is asymptotically efficient as there is a one-to-one relationship between the CCF and the transition density function. However, the practical implementation of the method involves two issues: namely the choice of the discrete grids at which the ECF and CCF are matched, and the weight function used in the estimation procedure. While in the univariate case the choice of discrete grids has received some attention, see Feuerverger and Mureika (1977), Schmidt (1982), Knight and Satchell (1997), Yu (1998) and Singleton (2001), in the multivariate case the choice is indeed an open question. With given discrete grids, Singleton (2001) proposes using a numerical procedure in the GMM framework to find the optimal weight function, and shows that the optimal weight function is simply the covariance of the ECF evaluated at the discrete grid points.

In this paper, we propose an alternative approach to address the issues involved in the ECF estimation procedure. Instead of choosing the discrete grid points first and then finding the optimal weight function following a numerical procedure, we propose a closed-form or analytical approximation of the optimal weight function in the ECF estimation procedure. As a result, the choice of the discrete grid can be avoided and the estimation can proceed based on the analytical conditional cumulants. The basic idea stems from the fact that, due to the one-to-one correspondence between the CCF and the transition density, the first-order conditions for ML estimation can indeed be written as the sum of a weighted integral of the difference between ECF and CCF. The optimal weight in this set-up is the inverse Fourier transform of the tth score. Thus by approximating the logarithmic transition density of the Markov process we can approximate the optimal weight function and hence solve the integral for the appropriate estimating equations. We demonstrate that the estimating equations are similar to the approximate maximum likelihood estimation (AMLE). However, in contrast to the conventional AMLE our approach ensures the consistency of the estimator even with the approximate optimal weight function.

The method applies to processes with closed-form CCF, e.g. the common continuous-time affine diffusion and jump-diffusion processes with known analytical CCF, the discrete time compound autoregressive (CAR) processes introduced by Darolles et al. (2006). The CAR processes are a class of non-linear dynamic processes characterized by the conditional log-Laplace transforms which are affine functions of lagged values of the process. Furthermore, since only the conditional cumulants are required, the method also applies to processes where the conditional cumulants to certain orders can be derived.

The paper is organized as follows. Section 2 proposes the ECF estimation approach with approximate optimal weight function and derives the appropriate estimating equations. In particular, we relate our approach to various existing estimation methods, especially the QMLE and AMLE. In Section 3, we illustrate the application of our approach with examples of various Markov processes in both discrete time and continuous time and for both univariate and multivariate cases. In Section 4, we perform Monte Carlo simulations to investigate the finite sample properties of the proposed estimation procedure. A brief conclusion is contained in Section 5. Proofs are given in Appendix A.

2. ECF METHOD AND CONSISTENT AMLE (C-AMLE)

Let $x_t \in \mathbb{R}^N$, $t > 0$, be an N-dimensional stationary Markov process defined in either a discrete time set or a continuous time set with a complete probability space $(\Omega, \mathcal{F}, P)$. Suppose that $\{x_t\}_{t=1}^T$ represents an observed sample over a discrete set of time from the Markov process. Let $f(x_{t+1} \mid x_t; \theta) : \mathbb{R}^N \times \mathbb{R}^N \times \Theta \to \mathbb{R}$ denote the measurable transition density for the Markov process and $\theta \in \Theta \subset \mathbb{R}^Q$ denote the parameter vector of the data-generating process for x_t. Following Singleton (2001), a consistent estimator of the parameters based on the empirical characteristic function (ECF) can be derived from the following equation:

$$\frac{1}{T}\sum_{t=1}^{T}\int .. \int w(r, t \mid x_t; \theta)(e^{ir'x_{t+1}} - \phi(r, x_{t+1} \mid x_t; \theta))\, dr = 0, \tag{2.1}$$

where $\phi(r, x_{t+1} \mid x_t; \theta) = E[\exp(ir'x_{t+1}) \mid x_t]$ is the CCF and $w(r, t \mid x_t; \theta) \in W(r, t \mid x_t; \theta)$ with $w(r, t \mid x_t; \theta)$ being the set of 'instrument' or 'weight' functions as defined in Singleton (2001). Namely, for each 'instrument' or 'weight' function $w(r, t \mid x_t; \theta) : \mathbb{R}^N \times \mathbb{R}^+ \times \mathbb{R}^N \times \Theta \to \mathbb{C}^Q$ where $\mathbb{C}^Q$ denotes the complex numbers, we have $w(r, t \mid x_t; \theta) \in I_t$ and $w(r, t \mid x_t; \theta) = \bar{w}(-r, t \mid x_t; \theta)$, $t = 1, 2, \ldots, T + 1$, where I_t is the σ-algebra generated by x_t. The ECF estimation procedure has been proposed by Feuerverger and Mureika (1977), Schmidt (1982) and Feuerverger and McDunnough (1981) for i.i.d. cases, and Feuerverger (1990) for generic stationary Markov processes using the joint ECF of state variables.

As shown in Feuerverger (1990), Singleton (2001), Jiang and Knight (2002) and Carrasco et al. (2007), there exists an optimal 'instrument' or 'weight' function in the sense that the estimator defined in (2.1) is also an efficient estimator with the same asymptotic properties as ML estimator. We summarize these results in the following lemma. Following Singleton (2001), we impose the same regularity conditions on the ECF estimation procedure as those imposed by Hansen (1982) for GMM.

LEMMA 2.1. *Under standard regularity conditions where $w(r, t \mid x_t; \theta) \in W(r, t \mid x_t; \theta)$ is a well-defined 'instrument' or 'weight' function, equation (2.1) leads to consistent parameter estimators. Furthermore, let $f(x_{t+1} \mid x_t; \theta)$ be the transition density of the Markov process, the following weight function is optimal:*

$$w(r, t \mid x_t; \theta) = \frac{1}{(2\pi)^N}\int .. \int \frac{\partial \ln f(x_{t+1} \mid x_t; \theta)}{\partial \theta} e^{-ir'x_{t+1}} dx_{t+1} \tag{2.2}$$

in the sense that the estimators based on equation (2.1) are equivalent to MLE.

It is noted that the optimal weight function is determined by the logarithm of the transition density or likelihood function of the Markov process. When $f(x_{t+1} \mid x_t; \theta)$ is explicitly known, the Markov model can be estimated straightforwardly via the ML method and it can be shown that the estimation equation (2.1) will result exactly in conditional ML estimator. However, if $f(x_{t+1} \mid x_t; \theta)$ is not known explicitly but $\phi(r, x_{t+1} \mid x_t; \theta)$ is, then (2.1) must be implemented with other than the optimal weight function $w(r, t \mid x_t; \theta)$.

Thus, the implementation of the estimation procedure in (2.1) involves two issues, namely the choice of the discrete grids at which the ECF and CCF are matched, and the weight function. Since with a discrete grid the estimation problem in (2.1) is essentially the GMM in the frequency domain, Singleton (2001) proposes a GMM approach to deal with the issues involved in the implementation of (2.1). Namely, the integral in (2.1) is first approximated with a sum over a discrete set of r and then the optimal weight function can be obtained by appealing to the GMM framework following a numerical procedure. As shown in Singleton (2001), under certain regularity conditions as the grid of r's becomes increasingly fine in $\mathbb{R}^N$, the asymptotic efficiency of the estimator approaches that of MLE. The appeal to GMM to find the optimal weight matrix in essence is a GLS solution but its major drawback is the necessity to choose the vectors r over which the sum and hence the integral is approximated. In the scalar Markov case the choice of the discrete grids has been considered in the literature, at least in the i.i.d. case; see Feuerverger and Mureika (1977), Schmidt (1982), Knight and Satchell (1997), Yu (1998) and Singleton (2001). In the N-dimensional case ($N \geq 2$) the problem is much more complicated and there is virtually no guidance given in the literature. In general, there is an obvious trade-off between finer grids and coarser grids. With too coarse a grid, the GMM estimation based on selected set of moments is easier to implement but would achieve lower asymptotic efficiency. While with too fine a grid, the implementation of the estimation procedure can become infeasible due to the singularity of the covariance matrix. Thus, in practice there is a limitation to the actual implementation of a very fine grid. In the context of the estimation of mixture distributions, Carrasco and Florens (2000) provide Monte Carlo evidence of efficiency loss relative to MLE due to the use of a discrete grid. In a recent paper, Carrasco et al. (2007) propose a GMM approach with a continuum of moment conditions and address the singularity problem via a penalization term.

In this paper, we propose an alternative approach than the one in Singleton (2001) to deal with the issues involved in the implementation of (2.1). Instead of choosing the vectors r first and then finding the optimal weight function following a numerical procedure, we propose an approximation to the optimal weight function in (2.2) and derive the estimating equations accordingly. As a result, the choice of the discrete grid is avoided and the estimation is relatively straightforward to implement. We show that our approach is similar to the approximate MLE as both are based on the approximate likelihood function or the approximate score. As in the case of the conventional approximate MLE, approximating the optimal weight function does result in an efficiency loss relative to MLE. However, in contrast to the conventional approximate MLE, our approach ensures the consistency of parameter estimators even with the approximated likelihood function.

2.1. Approximate optimal weight function and consistent AMLE (C-AMLE)

In this paper, we propose a closed-form or analytical approximation to the logarithmic transition density of the Markov process, i.e. $\ln f(x_{t+1} \mid x_t; \theta)$, which will then, via (2.2), give us an approximate weight function in (2.1) and hence result in a consistent estimator. We consider series expansions for the log transition density rather than for the density itself. In addition to the fact that the log transition density appears explicitly in the optimal weight function and thus is the function we aim to approximate, for a number of other reasons better approximations are often obtained by approximating the log transition density and then exponentiating. Since the solution of (2.1) requires the knowledge of $\phi(r, x_{t+1} \mid x_t; \theta)$ or $\ln \phi(r, x_{t+1} \mid x_t; \theta)$, we can use this function to develop approximations to $\ln f(x_{t+1} \mid x_t; \theta)$. The approximation we propose is the multivariate Edgeworth/Gram–Charlier expansion.

Following McCullagh (1987), using the tensor notation, we have the general Gram–Charlier/Edgeworth expansion for the log multivariate density $\ln f(x_{t+1} \mid x_t; \theta)$ given by

$$\begin{aligned}\ln f(x_{t+1} \mid x_t; \theta) = {} & \ln f_0(x_{t+1} \mid x_t; \theta) \\ & + \frac{1}{3!} K^{i,j,k} h_{ijk}(x_{t+1} \mid x_t) \\ & + \frac{1}{4!} K^{i,j,k,l} h_{ijkl}(x_{t+1} \mid x_t) \\ & + \frac{1}{5!} K^{i,j,k,l,m} h_{ijklm}(x_{t+1} \mid x_t) \\ & + \frac{1}{6!} [K^{i,j,k,l,m,n} h_{ijklmn}(x_{t+1} \mid x_t) + K^{i,j,k} K^{l,m,n} h_{ijk,lmn}(x_{t+1} \mid x_t)[10]] \\ & + \ldots, \end{aligned} \tag{2.3}$$

where $f_0(x_{t+1} \mid x_t; \theta)$ is chosen such that its first-order and second-order moments agree with those of x_{t+1} conditional on x_t. Upon letting $\psi(r, x_{t+1} \mid x_t; \theta) = \ln \phi(-ir, x_{t+1} \mid x_t; \theta)$ (the cumulant generating function), we have the conditional cumulants of various orders:

$$\begin{aligned} \lambda^i &= \frac{\partial}{\partial r_i} \psi(r, x_{t+1} \mid x_t; \theta)|_{r=0} \\ \lambda^{i,j} &= \frac{\partial^2}{\partial r_i \partial r_j} \psi(r, x_{t+1} | x_t; \theta)|_{r=0} \\ K^{i,j,k} &= \frac{\partial^3}{\partial r_i \partial r_j \partial r_k} \psi(r, x_{t+1} | x_t; \theta)|_{r=0} \\ K^{i,j,k,l} &= \frac{\partial^4}{\partial r_i \partial r_j \partial r_k \partial r_l} \psi(r, x_{t+1} | x_t; \theta)|_{r=0} \\ & \ldots . \end{aligned}$$

It is noted that Edgeworth series used for approximations to distributions are most conveniently expressed using cumulants. Moreover, where approximate normality is involved, higher-order cumulants can usually be neglected but not higher-order moments. In this paper, since we often deal with situations where the log CCF has simpler expression, the cumulants can be more conveniently obtained than the moments. Furthermore, the Hermite polynomial tensors in the general Edgeworth/Gram–Charlier expansion of equation (2.3) are given by

$$\begin{aligned} h_i &= \lambda_{i,j}(x^j - \lambda^j) \\ h_{ij} &= h_i h_j - \lambda_{i,j} \\ h_{ijk} &= h_i h_j h_k - h_i \lambda_{j,k}[3] \\ h_{ijkl} &= h_i h_j h_k h_l - h_i h_j \lambda_{k,l}[6] + \lambda_{i,j} \lambda_{k,l}[3] \\ h_{ijklm} &= h_i h_j h_k h_l h_m - h_i h_j h_k \lambda_{l,m}[10] + h_i \lambda_{j,k} \lambda_{l,m}[15] \end{aligned}$$

$$h_{ijklmn} = h_i \ldots h_n - h_i h_j h_k h_l \lambda_{m,n}[15] + h_i h_j \lambda_{k,l}\lambda_{m,n}[45] - \lambda_{i,j}\lambda_{k,l}\lambda_{m,n}[15]$$
$$\ldots$$
$$h_{ijk,lmn} = h_{ijklmn} - h_{ijk}h_{lmn}.$$

In tensor notation, it is understood that any index repeated once as a subscript and once as a superscript is interpreted as sums over these repeated scripts, i.e. $h_i = \lambda_{i,j}(x^j - \lambda^j) = \sum_j \lambda_{i,j}(x^j - \lambda^j)$, etc. Also, the numbers in square brackets refer to the number of permutations of the various subscripts.

In this paper, for simplicity we let the initial approximating function be the multivariate normal density with mean vector λ^i and covariance matrix $\lambda^{i,j}$, i.e.

$$f_0(x_{t+1} \mid x_t; \theta) = (2\pi)^{-N/2}|\lambda^{i,j}|^{-1/2} \exp\left(-\frac{1}{2}\left(x_{t+1}^i - \lambda^i\right)\left(x_{t+1}^j - \lambda^j\right)\lambda_{i,j}\right), \tag{2.4}$$

with x_{t+1} being an $N \times 1$ vector whose ith element is x_{t+1}^i, mean λ^i and covariance matrix $\lambda^{i,j}$, $\lambda_{i,j}$ is the inverse matrix of $\lambda^{i,j}$ and $|\lambda^{i,j}|$ is the determinant of the covariance matrix.

For clarity of presentation as well as ease of notation and yet without loss of generality, in the following discussion we focus on the Gram–Charlier series expansion and set the truncation order $p = 4$; consequently we have,

$$\begin{aligned} \ln \hat{f}_p(x_{t+1} \mid x_t; \theta) &= -\frac{N}{2}\ln 2\pi - \frac{1}{2}\ln|\lambda^{i,j}| - \frac{1}{2}(x^i - \lambda^i)(x^j - \lambda^j)\lambda_{i,j} \\ &\quad + \frac{1}{3!}K^{i,j,k}h_{ijk} + \frac{1}{4!}K^{i,j,k,l}h_{ijkl} \\ &= \ln f(x_{t+1} \mid x_t; \theta) - \ln f_p^{\Delta}(x_{t+1} \mid x_t; \theta), \end{aligned} \tag{2.5}$$

where $\ln f_p^{\Delta}(x_{t+1} \mid x_t; \theta)$ is the approximation error. The Edgeworth series and Gram–Charlier series are formally identical when the expansion order is infinite and the main difference is the different criteria used in collecting terms in a truncated series. As a result, with the same order of expansion for the Edgeworth series and Gram–Charlier series, different cumulants or moments may appear in the estimating equations.

Let the parameter vector to be estimated be denoted by $\theta \in \Theta$, then all cumulants and the Hermite tensors are functions of θ. The approximate score function, i.e. the derivative of the $\ln \hat{f}_p(x_{t+1} \mid x_t; \theta)$, is given by

$$\begin{aligned} \frac{\partial \ln \hat{f}_p(x_{t+1} \mid x_t; \theta)}{\partial \theta} &= -\frac{1}{2|\lambda^{i,j}|}\frac{\partial}{\partial \theta}|\lambda^{i,j}| + \frac{\partial \lambda^i}{\partial \theta}h_i - \frac{1}{2}(x^i - \lambda^i)(x^j - \lambda^j)\frac{\partial \lambda_{i,j}}{\partial \theta} \\ &\quad + \frac{1}{6}\left[\frac{\partial K^{i,j,k}}{\partial \theta}h_{ijk} + K^{i,j,k}\frac{\partial h_{ijk}}{\partial \theta}\right] \\ &\quad + \frac{1}{24}\left[\frac{\partial K^{i,j,k,l}}{\partial \theta}h_{ijkl} + K^{i,j,k,l}\frac{\partial h_{ijkl}}{\partial \theta}\right]. \end{aligned} \tag{2.6}$$

Using the approximate score in (2.6), we can define an approximate optimal weight function from (2.2) as

$$\hat{\omega}_p(r, t \mid x_t; \theta) = \frac{1}{(2\pi)^N}\int \ldots \int \frac{\partial \ln \hat{f}_p(x_{t+1} \mid x_t; \theta)}{\partial \theta} e^{-ir'x_{t+1}} dx_{t+1}. \tag{2.7}$$

The following theorem states the estimating equation for a consistent estimator of θ. Since the estimation is based on an approximate likelihood function, we refer to the estimator as the AMLE. The asymptotic distribution of the proposed estimator is derived under standard regularity conditions as in Singleton (2001).

THEOREM 2.1. *(Consistent Approximate MLE) Given the approximate optimal weight function in (2.7), the ECF procedure in (2.1) can be written as*

$$\frac{1}{T}\sum_{t=1}^{T}\int \hat{\omega}_p(r,t\,|\,x_t;\theta)(e^{ir'x_{t+1}} - \phi(r,x_{t+1}\,|\,x_t;\theta))dr = 0, \tag{2.8}$$

with $w(r,t\,|\,x_t;\theta)$ replaced by $\hat{\omega}_p(r,t\,|\,x_t;\theta)$. From the above equation, we have the following estimating equation for our proposed estimator which we refer to as the approximate MLE. That is

$$\frac{1}{T}\sum_{t=1}^{T}\left\{\frac{\partial \ln \hat{f}_p(x_{t+1}\,|\,x_t;\theta)}{\partial\theta} - E\left[\frac{\partial \ln \hat{f}_p(x_{t+1}\,|\,x_t;\theta)}{\partial\theta}\,\middle|\,x_t\right]\right\} = 0. \tag{2.9}$$

Under standard regularity conditions, the approximate ML estimator defined in (2.8) or (2.9) is consistent and asymptotically normal, i.e. as $T\to\infty$:

$$\sqrt{T}(\hat{\theta}_p - \theta) \xrightarrow{d} N(0,\Omega_p), \tag{2.10}$$

where the limiting covariance matrix Ω_p is given in Appendix A.

The estimating equation in (2.9) is based on the approximate likelihood function or the approximate score.[1] However, in contrast to the conventional AMLE it also includes the second term, i.e. the expectation of the approximate score. It is clear that when the true transition density of the Markov process $f(x_{t+1}\,|\,x_t;\theta)$ is known and $\hat{f}_p(x_{t+1}\,|\,x_t;\theta) = f(x_{t+1}\,|\,x_t;\theta)$, then we have $E[\frac{\partial \ln \hat{f}_p(x_{t+1}\,|\,x_t;\theta)}{\partial\theta}] = 0$. However, this will not necessarily be the case if we approximate $f(x_{t+1}\,|\,x_t;\theta)$, i.e. $\hat{f}_p(x_{t+1}\,|\,x_t;\theta) \neq f(x_{t+1}\,|\,x_t;\theta)$. The inclusion of the second term ensures the consistency of the proposed estimator and it does not require convergence of the infinite expansion. However, as noted in the proof, if as p goes to infinity we have $\ln \hat{f}_p \to \ln f$, i.e. convergence, then the estimator becomes the ECF estimator proposed in Singleton (2001) with optimal weight function and thus achieves ML efficiency. Conditions needed for the convergence of the Edgeworth/Gram–Charlier series are given in Cramér (1925) for the univariate case and in Skovgaard (1986) for the multivariate case. As Cramér (1946, p. 224) notes, 'in practical applications it is in most cases only of little value to know the convergence properties of our expansions. What we really want to know is whether a small number of terms—usually not more than two or three—suffice to give a good approximation to $f(x)$'. In our simulations, reported in Section 4, we note that an expansion involving corrections for skewness and kurtosis does indeed work very well.

[1] Note that the estimating equation in (2.9) is equivalent to the ECF procedure in (2.8). One can thus resort to (2.8) for model estimation when the conditional expectation in (2.9) is difficult to compute. We note that when analytical expression of the CCF is unavailable, an approximate efficient estimator based on discretizing the integral in (2.8) is proposed in Singleton (2001) and a simulated method of moments is proposed in Carrasco et al. (2007).

Since the QML estimation is a special case of approximate MLE, it is also a special case of our estimator (with expansion $p = 2$). In fact, the consistency of the QML estimation can be easily seen from our framework. For this and one technical note, we add the following remarks.

REMARK 2.1. If one was merely to approximate $f(x_{t+1} \mid x_t; \theta)$ by a normal density, i.e. only the initial approximating density $\hat{f}_p(x_{t+1} \mid x_t; \theta) = f_0(x_{t+1} \mid x_t; \theta)$ and $\hat{f}_p(x_{t+1} \mid x_t; \theta) \neq f(x_{t+1} \mid x_t; \theta)$, the approach is essentially the QML estimation. From (2.4), we have

$$E\left[\frac{\partial \ln \hat{f}_p(x_{t+1} \mid x_t\ ; \theta)}{\partial \theta}\right] = 0$$

which shows, via the results in Bollerslev and Wooldridge (1992), the consistency of QML estimation.

REMARK 2.2. The reader may be concerned that in the definitions (2.2) and (2.7) the score and approximate score may not be integrable. However, using the theory of generalized functions these integrals will exist. In the univariate case, the approximate score is a polynomial in x_{t+1} and consequently the approximate optimal weight will be a function involving the Dirac delta function and its derivatives. The result is similar in the multivariate case with the multivariate Dirac delta function playing the central role.

2.2. *The approximate MLE (C-AMLE) versus other methods*

In the following we further derive the estimating equations based on (2.9) and illustrate the similarity and difference of our estimator with other estimation methods. From the definition of the Hermite polynomials we can readily establish that

$$\begin{aligned}
\frac{\partial h_i}{\partial \theta} &= \frac{\partial \lambda_{i,j}}{\partial \theta}(x^j - \lambda^j) - \lambda_{i,j}\frac{\partial \lambda^j}{\partial \theta} \\
&= \bar{\lambda}_{i,j}(x^j - \lambda^j) - \lambda_{i,j}\bar{\lambda}^j = \bar{h}_i - \bar{z}_i \\
\frac{\partial h_{ijk}}{\partial \theta} &= \frac{\partial h_i}{\partial \theta} h_j h_k[3] - \frac{\partial h_i}{\partial \theta}\lambda_{j,k}[3] - h_i \frac{\partial \lambda_{j,k}}{\partial \theta}[3] \\
\frac{\partial h_{ijkl}}{\partial \theta} &= \frac{\partial h_i}{\partial \theta} h_j h_k h_l[4] - h_i h_j \frac{\partial \lambda_{k,l}}{\partial \theta}[6] - \frac{\partial h_i}{\partial \theta} h_j \lambda_{k,l}[12] + \frac{\partial \lambda_{i,j}}{\partial \theta}\lambda_{k,l}[6].
\end{aligned}$$

Substituting these derivatives into the expansion given by (2.6) and taking expectations we can derive the appropriate estimating equations, which are stated in the following lemma.

LEMMA 2.2. *For an N-dimensional Markov process with known CCF associated with an unknown transition density, following the ECF estimation procedure with approximate optimal weight function the use of an Edgeworth/Gram–Charlier approximation for the unknown* *transition density as in (2.5) results in the following estimating equations.*

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\Bigg\{\frac{\partial\lambda^i}{\partial\theta}h_i-\frac{1}{2}[(x^i-\lambda^i)(x^j-\lambda^j)-\lambda^{i,j}]\frac{\partial\lambda_{i,j}}{\partial\theta}\\
&+\frac{1}{6}\left[\frac{\partial K^{i,j,k}}{\partial\theta}(h_{ijk}-E(h_{ijk}\,|\,x_t))+3K^{i,j,k}\left[(\bar{h}_ih_jh_k-E[\bar{h}_ih_jh_k\,|\,x_t])\right.\right.\\
&\left.\left.-\bar{z}_i(h_jh_k-E[h_jh_k\,|\,x_t])-\bar{h}_i\,\lambda_{j,k}-h_i\,\frac{\partial\lambda_{j,k}}{\partial\theta}\right]\right]\\
&+\frac{1}{24}\left[\frac{\partial K^{i,j,k,l}}{\partial\theta}(h_{ijkl}-E(h_{ijkl}\,|\,x_t))+K^{i,j,k,l}\left[4(\bar{h}_ih_jh_kh_l-E[\bar{h}_ih_jh_kh_l\,|\,x_t])\right.\right.\\
&-4\bar{z}_i(h_jh_kh_l-E[h_jh_kh_l\,|\,x_t])-12(\bar{h}_ih_j\lambda_{k,l}-E[\bar{h}_ih_j\lambda_{k,l}\,|\,x_t])\\
&\left.\left.+12\bar{z}_ih_j\lambda_{k,l}-6(h_ih_j-E[h_ih_j\,|\,x_t])\frac{\partial\lambda_{k,l}}{\partial\theta}\right]\right]\Bigg\}=0.
\end{aligned}
\tag{2.11}
$$

If θ is of dimension Q then there will be Q such equations, the solution of which will lead to approximate ML estimation.

The results in Lemma 2.2 underline the distinguishing feature of our method, i.e. while the estimating equations are derived based on the approximation of scores, the approximate scores are not used directly for estimation as in the common AMLE. Instead, these scores are used to construct the optimal weight function in the ECF estimation procedure and the estimating equations only involve certain conditional cumulants. Clearly, an advantage of the proposed method is that since only the conditional cumulants are required in the estimation, the method applies to general Markov processes where the conditional cumulants to certain orders are available in closed form. The following remark discusses the relationship between our method and alternative estimation methods in the literature.

REMARK 2.3. The estimating equations in (2.11) show clearly that similar to the method of moments (MM) and the GMM, our method is also based on the conditional cumulants or equivalently conditional moments of various orders. While our method is similar to the MM in that the number of moment restrictions is the same as the dimension of the parameter vector, our estimating equations may involve more moments than the dimension of the parameter vector. While our method is similar to the GMM in that the number of moments involved in the estimation may be higher than the dimension of the parameter vector, the number of moment restrictions is the same as the dimension of the parameter vector. In particular, GMM relies on a numerical procedure to find the optimal weights for various moment conditions, the moment restrictions in our method have their own non-linear structure or specific weights of various moment conditions as defined in the estimating equations.

While the estimating equations in the general case are cumbersome as in (2.11), in the univariate case they collapse into the well-known method of moments as we will now illustrate, albeit the moment restrictions are a system of conventional moment conditions. In the univariate case we essentially just let $i = j = k = l = 1$ and thus drop the superscript index

$$\begin{aligned}
&\lambda_{11} = 1/K_2 \\
&h_1 = (x_{t+1} - K_1)/K_2, \quad h_{11} = h_2 = h_1^2 - 1/K_2 \\
&h_{111} = h_1^3 - 3h_1/K_2, \quad h_{1111} = h_1^4 - 6h_1^2/K_2 + 3/K_2^2
\end{aligned}$$

with

$$\begin{aligned}
\frac{\partial h_1}{\partial\theta} &= -h_1\frac{\partial K_2/\partial\theta}{K_2} - \frac{\partial K_1/\partial\theta}{K_2}, &\quad h_1\frac{\partial h_1}{\partial\theta} &= -h_1^2\frac{\partial K_2/\partial\theta}{K_2} - h_1\frac{\partial K_1/\partial\theta}{K_2} \\
h_1^2\frac{\partial h_1}{\partial\theta} &= -h_1^3\frac{\partial K_2/\partial\theta}{K_2} - h_1^2\frac{\partial K_1/\partial\theta}{K_2}, &\quad h_1^3\frac{\partial h_1}{\partial\theta} &= -h_1^4\frac{\partial K_2/\partial\theta}{K_2} - h_1^3\frac{\partial K_1/\partial\theta}{K_2}
\end{aligned}$$

and since

$$\begin{aligned}
&E[h_1 \,|\, x_t] = 0, \quad E\big[h_1^2 \,|\, x_t\big] = 1/K_2 \\
&E[h_{11} \,|\, x_t] = 0, \quad E\big[h_{111} \,|\, x_t\big] = K_3/K_2^3 \\
&E[h_{1111} \,|\, x_t] = E\big[h_1^4 \,|\, x_t\big] - 6E\big[h_1^2 \,|\, x_t\big]/K_2 + 3/K_2^2 = K_4/K_2^4.
\end{aligned}$$

The estimating equations collapse to

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} h_1\frac{\partial K_1}{\partial\theta} &+ \frac{1}{2}\left(h_1^2 - \frac{1}{K_2}\right)\frac{\partial K_2}{\partial\theta} + \frac{1}{6}\left[\frac{\partial K_3}{\partial\theta}\left(h_1^3 - \frac{K_3}{K_2^3}\right) - \frac{3\partial K_3/\partial\theta}{K_2}h_1\right] \\
&- \frac{K_3}{2}\left[\left(h_1^3 - \frac{K_3}{K_2^3}\right)\frac{\partial K_2/\partial\theta}{K_2} + \left(h_1^2 - \frac{1}{K_2}\right)\frac{\partial K_1/\partial\theta}{K_2} - 2h_1\frac{\partial K_2/\partial\theta}{K_2^2}\right] \\
&+ \frac{1}{24}\left[\frac{\partial K_4}{\partial\theta}\left(h_1^4 - \frac{(K_4 + 3K_2^2)}{K_2^4}\right) - 6\frac{\partial K_4/\partial\theta}{K_2}\left(h_1^2 - \frac{1}{K_2}\right)\right] \\
&- \frac{K_4}{24}\left[4\left(h_1^4 - \left(\frac{K_4 + 3K_2^2}{K_2^4}\right)\right)\frac{\partial K_2/\partial\theta}{K_2} + 4\left(h_1^3 - \frac{K_3}{K_2^3}\right)\frac{\partial K_1/\partial\theta}{K_2}\right. \\
&\left. - 18\left(h_1^2 - \frac{1}{K_2}\right)\frac{\partial K_2/\partial\theta}{K^2} - 12h_1\frac{\partial K_1/\partial\theta}{K_2^2}\right] = 0.
\end{aligned} \tag{2.12}$$

Again the derivatives are taken with respect to all elements in the Q-dimensional parameter vector θ.

The above estimating equations can be readily put into a more recognizable form by combining coefficients on $(h_1^j - E(h_1^j \,|\, x_t))$, $j = 1, 2, 3, 4$ $(p = 4)$. Letting A_{jt}^i be the appropriate coefficient on $(h_1^j - E(h_1^j \,|\, x_t))$, $j = 1, 2, 3, 4$ $(p = 4)$, associated with the derivative with respect to the ith element of θ, we have the estimating equations given by

$$\frac{1}{T}\sum_{t=1}^{T} A_t g_t = 0,$$

where A_t is a $Q \times 4$ ($p = 4$) matrix with the ith row being associated with $\frac{\partial}{\partial \theta^i}$ and g_t is a 4×1 ($p = 4$) vector given by

$$g_t = \begin{bmatrix} h_1 \\ h_1^2 - 1/K_2 \\ h_1^3 - K_3/K_2^3 \\ h_1^4 - \left(K_4 + 3K_2^2\right)/K_2^4 \end{bmatrix}.$$

More specifically, we have

$$\begin{aligned}
A_{1t} &= \frac{\partial K_1}{\partial \theta} - \frac{\partial K_3/\partial \theta}{2K_2} + K_3 \frac{\partial K_2/\partial \theta}{K_2^2} + K_4 \frac{\partial K_1/\partial \theta}{2K_2^2} \\
A_{2t} &= \frac{\partial K_2/\partial \theta}{2} - K_3 \frac{\partial K_1/\partial \theta}{2K_2} - \frac{\partial K_4/\partial \theta}{4K_2} + 3K_4 \frac{\partial K_2/\partial \theta}{4K_2^2} \\
A_{3t} &= \frac{\partial K_3/\partial \theta}{6} - K_3 \frac{\partial K_2/\partial \theta}{2K_2} - K_4 \frac{\partial K_1/\partial \theta}{6K_2} \\
A_{4t} &= \frac{\partial K_4/\partial \theta}{24} - K_4 \frac{\partial K_2/\partial \theta}{6K_2}.
\end{aligned}$$

3. ILLUSTRATIVE EXAMPLES OF MARKOV PROCESSES

EXAMPLE 3.1. The Ornstein–Uhlenbeck Process (Equivalence to MLE) (Univariate Continuous-Time Gaussian Process). The OU process is a univariate diffusion process specified by the following stochastic differential equation:

$$dx_t = \beta(\alpha - x_t)dt + \sigma dw_t, \tag{3.1}$$

where w_t is a standard Brownian motion. The OU process has a normal transition density function given by $f(x_{t+\tau} \mid x_t; \theta) = \frac{1}{\sqrt{2\pi s^2}} \exp\{-\frac{(x_{t+\tau} - \alpha - (x_t - \alpha)e^{-\beta\tau})^2}{2s^2}\}$, where $s^2 = \frac{\sigma^2}{2\beta}(1 - e^{-2\beta\tau})$. The conditional log-likelihood is given by $\ln L = \sum_{t=1}^{T} \ln f(x_{t+1} \mid x_t; \theta)$ and maximization of the likelihood function leads to the ML estimator.

As a member of the affine class of diffusions, the CCF of the OU process is given by $\phi(r, x_{t+1} \mid x_t; \theta) = \exp\{ir(\alpha + (x_t - \alpha)e^{-\beta}) - \frac{r^2\sigma^2}{4\beta}(1 - e^{-2\beta})\}$. The conditional cumulants can be easily derived as

$$\begin{aligned}
K_1 &= \left(\alpha + (x_t - \alpha)e^{-\beta}\right) \\
K_2 &= \frac{\sigma^2}{2\beta}(1 - e^{-2\beta}), \quad K_i = 0, \forall\, i \geq 3.
\end{aligned}$$

Substituting the cumulants into the estimating equation in (2.12) for the univariate case, we have

$$\frac{1}{T} \sum_{t=1}^{T} \left(h_1 \frac{\partial K_1}{\partial \theta} + \frac{1}{2} \left(h_1^2 - \frac{1}{K_2} \right) \frac{\partial K_2}{\partial \theta} \right) = 0,$$

where $h_1 = (x_{t+1} - K_1)/K_2$, $\theta = (\alpha, \beta, \sigma)$. It is straightforward to verify that this is equivalent to MLE.

EXAMPLE 3.2. The Square-Root Diffusion Process (Univariate Continuous-Time Non-Gaussian Process). The square-root process is a univariate diffusion specified by the following stochastic differential equation:

$$dx_t = \beta(\alpha - x_t)dt + \sigma\sqrt{x_t}dw_t. \tag{3.2}$$

The transition density of the square-root process is non-central chi-square and the marginal density is a gamma function. The square-root process is also a member of the affine class of diffusions and has the following CCF: $\phi(r, x_{t+\tau} \mid x_t; \theta) = (1 - \frac{ir}{c})^{-(q+1)} \exp\{\frac{ire^{-\beta\tau}}{(1-ir/c)}x_t\}$, where $c = 2\beta/(\sigma^2(1 - e^{-\beta\tau}))$, $q = \frac{2\alpha\beta}{\sigma^2} - 1$. The following four cumulants can be easily derived:

$$\begin{aligned}
K_1 &= \alpha(1 - e^{-\beta\tau}) + x_t e^{-\beta\tau} \\
K_2 &= \frac{\alpha\sigma^2}{2\beta}(1 - e^{-\beta\tau})^2 + \frac{x_t\sigma^2}{\beta}e^{-\beta\tau}(1 - e^{-\beta\tau}) \\
K_3 &= \frac{\alpha\sigma^4}{2\beta^2}(1 - e^{-\beta\tau})^3 + \frac{3x_t\sigma^4}{2\beta^2}e^{-\beta\tau}(1 - e^{-\beta\tau})^2 \\
K_4 &= \frac{3\alpha\sigma^6}{4\beta^3}(1 - e^{-\beta\tau})^4 + \frac{3x_t\sigma^6}{\beta^3}e^{-\beta\tau}(1 - e^{-\beta\tau})^3.
\end{aligned}$$

Again, it is straightforward to apply the estimating equation in (2.12) for the univariate case using the above results. The derivatives of these cumulants with respect to the parameter vector $\theta = (\alpha, \beta, \sigma^2)$ can be easily derived.

EXAMPLE 3.3. The Autoregressive Gamma Process (Univariate Discrete-Time Non-Linear Non-Gaussian Process). Darolles et al. (2006) introduced a class of non-linear Markov processes, termed CAR processes, in the following general non-linear Markov framework:

$$y_t = h(y_{t-1}, \epsilon_t, \theta),$$

where ϵ_t is an i.i.d. stochastic process and θ is the parameter vector. The CAR processes are characterized by the conditional log-Laplace transforms which are affine functions of the lagged variables of the process. With the closed-form conditional log-Laplace transforms or equivalently the CCF of the state variables, the method proposed in this paper can be readily used for the statistical inference of this class of processes.

Among the CAR processes, a very interesting model is the discrete time counterpart of the square-root diffusion process (as presented in Example 3.2 above) in continuous time, namely the autoregressive gamma process. The process is specified through the following compounding distribution:

$$\begin{aligned}
y_t \mid x_t &\sim \gamma(\delta + x_t) \\
x_t \mid y_{t-1} &\sim \mathcal{P}(\beta y_{t-1}),
\end{aligned} \tag{3.3}$$

where $\gamma(\cdot)$ and $\mathcal{P}(\cdot)$ are the gamma and Poisson distributions, respectively, and $\theta = (\beta, \delta)$ is the parameter vector. Similar to the square-root diffusion process in continuous-time as in Example 3.2, the transition density of the process has a non-central gamma distribution and the marginal density of the process has a gamma distribution, namely, $y_t \mid y_{t-1} \sim \gamma(\delta, \beta y_{t-1})$ and $(1 - \beta)y_t \sim \gamma(\delta)$. The process first introduced in Gouriéroux and Jasiak (2000) has the following CCF: $\phi(r, y_t \mid y_{t-1}) = E[\exp\{iry_t \mid y_{t-1}\}] = \exp\{-a(r)y_{t-1} + b(r)\}$, where $a(r) = -i\beta r/(1 - ir)$, $b(r) = -\delta \ln(1 - ir)$ and the marginal characteristic function $\psi(r, y_t) = E[\exp\{iry_t\}] =$

$\exp\{c(r)\}$, where $c(r) = -\delta \ln(1 - ir/(1-\beta))$. The process is stationary when $|\beta| < 1$. The conditional cumulants can be derived as

$$K_1 = \beta y_t + \delta, \qquad K_2 = 2\beta y_t + \delta$$
$$K_3 = 6\beta y_t + 2\delta, \qquad K_4 = 24\beta y_t + 6\delta.$$

It is straightforward to apply the estimating equation in (2.12) for the univariate case using the above results.

EXAMPLE 3.4. The Double OU Process (Equivalence to MLE) (The Bivariate Continuous-Time Gaussian Process)

$$\begin{aligned} dy_t &= \kappa(x_t - y_t)dt + \sigma dw_t^y \\ dx_t &= -\beta x_t dt + \sigma_x dw_t^x, \quad t \in [0, T], \end{aligned} \tag{3.4}$$

where $\kappa, \beta > 0, \kappa \neq \beta$. The process y_t exhibits linear mean reversion to conditional mean x_t which itself follows a mean-reverting process. The transition density of the process is given by

$$f(y_{t+\tau}, x_{t+\tau} \mid \mathcal{F}_t) = (2\pi)^{-1}|V|^{-1/2} \cdot \exp\left\{-\frac{1}{2}\begin{pmatrix} y_{t+\tau} - m_y \\ x_{t+\tau} - m_x \end{pmatrix}' V^{-1} \begin{pmatrix} y_{t+\tau} - m_y \\ x_{t+\tau} - m_x \end{pmatrix}\right\},$$

where $m_x = E[x_{t+\tau} \mid \mathcal{F}_t]$, $m_y = E[y_{t+\tau} \mid \mathcal{F}_t]$ and

$$V = \begin{pmatrix} \mathrm{Var}[y_{t+\tau} \mid \mathcal{F}_t] & \mathrm{Cov}(y_{t+\tau}, x_{t+\tau} \mid \mathcal{F}_t) \\ \mathrm{Cov}(x_{t+\tau}, y_{t+\tau} \mid \mathcal{F}_t) & \mathrm{Var}[x_{t+\tau} \mid \mathcal{F}_t] \end{pmatrix}.$$

The conditional likelihood is given by $\ln L = \sum_{t=1}^{T} \ln f(y_{t+1}, x_{t+1} \mid y_t, x_t; \theta)$, where $\theta = (\kappa, \sigma, \beta, \sigma_x)$ and maximizing the likelihood function leads to the ML estimator.

The joint CCF of $(y_{t+\tau}, x_{t+\tau})$ can be derived in closed form and is given in Appendix B. The conditional cumulants of various orders can be derived either from the closed-form expression of CCF or through the moment–cumulant relationship:

$$\begin{aligned} \mu_1 &= K_1 \\ \mu_2 &= K_2 + K_1^2 \\ \mu_3 &= K_3 + 3K_2K_1 + K_1^3 \\ \mu_4 &= K_4 + 4K_3K_1 + 3K_2^2 + 6K_2K_1^2 + K_1^4. \end{aligned}$$

For relations between higher-order moments and cumulants, see Kendall and Stuart (1977). Note that the superscript notation is used for cumulants in the multivariate case and 1 refers to the cumulants associated with y_t and 2 refers to the cumulants associated with x_t. We have

$$K^1 = E[y_{t+\tau} \mid \mathcal{F}_t], \qquad K^{1,1} = \mathrm{Var}[y_{t+\tau} \mid \mathcal{F}_t], \qquad K^{1,1,1} = K^{1,1,1,1} = 0$$
$$K^2 = E[x_{t+\tau} \mid \mathcal{F}_t], \qquad K^{2,2} = \mathrm{Var}[x_{t+\tau} \mid \mathcal{F}_t], \qquad K^{2,2,2} = K^{2,2,2,2} = 0$$
$$K^{1,2} = \mathrm{Cov}(y_{t+\tau}, x_{t+\tau} \mid \mathcal{F}_t), \qquad K^{1,1,2} = K^{1,2,2} = K^{1,1,2,2} = 0.$$

Substituting the cumulants into the estimating equation in (2.11) for the multivariate case, we have the jth estimating equation given by

$$\frac{1}{T}\sum_{t=1}^{T}\left\{\frac{\partial K'}{\partial\theta_j}h+\frac{1}{2}\left(h'\frac{\partial V}{\partial\theta_j}h-tr\left(\frac{\partial V}{\partial\theta_j}V^{-1}\right)\right)\right\}=0,$$

where $h'=(h_1,h_2)$, $\frac{\partial K'}{\partial\theta_j}=(\frac{\partial K^1}{\partial\theta_j},\frac{\partial K^2}{\partial\theta_j})$, and here we have used the fact that $\frac{\partial V^{-1}}{\partial\theta_j}=-V^{-1}\frac{\partial V}{\partial\theta_j}V^{-1}$, with $j=1,2,3,4$. It is straightforward to verify that this is equivalent to ML estimation.

EXAMPLE 3.5. The Discrete Time Stochastic Volatility Process (Bivariate Discrete Time Non-Linear Process). The following process is often used in the finance literature to model the asset return process with stochastic conditional volatility.

$$\begin{aligned}x_t&=e^{h_t/2}\epsilon_t\\ h_t&=\alpha+\beta h_{t-1}+\sigma\nu_t\\ \begin{pmatrix}\epsilon_t\\ \nu_t\end{pmatrix}&\sim N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1&\rho\\ \rho&1\end{pmatrix}\right).\end{aligned}\tag{3.5}$$

It is noted that the return process and stochastic volatility are correlated, reflecting the leverage effect when $\rho<0$. While the stochastic volatility is in general unobserved, a set of recent literature has shown that it can be either estimated from high frequency return observations or implied from the derivatives market. As demonstrated in recent studies such as Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002a,b), daily volatility can be very well approximated from high frequency intradaily returns. Such an estimator or proxy of stochastic volatility can be readily used for the estimation of the process, and in the subsequent analysis we apply the proposed method to the SV process in the context of bivariate Markov processes.

For the model specified in (3.5) there is no closed-form expression for the CCF; however, we can still derive the necessary conditional moments of x_t,h_t and then use them in our estimation. Substituting h_t into the process of x_t, we have

$$x_t=e^{(\alpha+\beta h_{t-1})/2}\cdot e^{\sigma\nu_t/2}\cdot\epsilon_t.$$

Let $g_t=h_t-\alpha-\beta h_{t-1}$, we aim to derive the conditional moments $E[x_t^n g_t^m\mid h_{t-1}],n,m=0,1,\ldots$. Decomposing the correlated normal random variables, we have

$$\begin{aligned}E\big[x_t^n g_t^m\mid h_{t-1}\big]&=e^{n(\alpha+\beta h_{t-1})/2}\cdot E\big[e^{n\sigma\sqrt{1-\rho^2}u_t/2}\cdot e^{n\sigma\rho\epsilon_t/2}\cdot\epsilon_t^n\cdot(\sigma\nu_t)^m\mid h_t\big]\\ &=e^{n(\alpha+\beta h_{t-1})/2}\cdot\sum_{q=0}^{m}\begin{pmatrix}m\\q\end{pmatrix}E\big[\big(\sigma\sqrt{1-\rho^2}u_t\big)^{m-q}e^{n\sigma\sqrt{1-\rho^2}u_t/2}\big]\\ &\quad\cdot E[(\sigma\rho\epsilon_t)^q\epsilon_t^n e^{n\sigma\rho\epsilon_t/2}].\end{aligned}$$

Since $E[z^n e^{rz}]=\varphi^{(n)}(r)=\frac{\partial^n e^{r^2/2}}{\partial r^n}$, thus,

$$\begin{aligned}E\big[\big(\sigma\sqrt{1-\rho^2}u_t\big)^{m-q}e^{n\sigma\sqrt{1-\rho^2}u_t/2}\big]&=\big(\sigma\sqrt{1-\rho^2}\big)^{m-q}\varphi^{(m-q)}\big(n\sigma\sqrt{1-\rho^2}/2\big)\\ E\big[(\sigma\rho\epsilon_t)^q\epsilon_t^n e^{n\sigma\rho\epsilon_t/2}\big]&=(\sigma\rho)^q\varphi^{(q+n)}(n\sigma\rho/2).\end{aligned}$$

The conditional moments of various orders can be readily derived. For instance, when $n = m = 2$, we have $E[x_t^2 g_t^2 \mid h_{t-1}] = \sigma^2(1+\sigma^2+\rho^2(2+5\sigma^2+\sigma^4)) \cdot e^{\alpha+\beta h_{t-1}+\sigma^2/2}$. The conditional cumulants of various orders can be easily derived via the cumulant–moment relationship and substituting the cumulants into (2.11) for the multivariate case, we have the necessary estimating equations.

EXAMPLE 3.6. The Continuous-Time Square-Root Volatility Process (Bivariate Continuous-Time Process). The following process with stochastic volatility is often used in the literature to model asset return in the continuous time framework:

$$\begin{aligned} dx_t &= \mu dt + \sqrt{v_t} dw_t^s \\ dv_t &= \beta(\alpha - v_t) dt + \sigma\sqrt{v_t} dw_t^v \\ dw_t^s dw_t^v &= \rho dt, \quad t \in [0, T], \end{aligned} \tag{3.6}$$

where $\beta > 0$. The process is known to be a positive process with a reflecting barrier at 0, which is attainable when $2\alpha\beta < \sigma^2$. This model has been widely used in the finance literature for asset return dynamics as it has an associated closed-form expression for European option prices; see Heston (1993). Similar to the discrete-time SV model, we apply the proposed method in this paper to the continuous-time SV process in the context of bivariate Markov processes.

The joint CCF of $(x_{t+\tau}, v_{t+\tau})$ can be derived in closed from and is given in Appendix B. Based on the closed-form expression of the characteristic function, analytical expressions for the conditional cumulants of any order can be derived as follows with 1 for x_t and 2 for v_t, respectively.

$$\begin{aligned}
K^1 &= \mu\tau \\
K^{1,1} &= \alpha\tau + \alpha\frac{1-e^{\beta\tau}}{\beta e^{k\beta\tau}} + \frac{e^{\beta\tau}-1}{\beta e^{\beta k\tau}} v_t \\
K^{1,1,1} &= \frac{3e^{-\beta\tau}}{\beta^2}(\alpha(2+\beta\tau-e^{\beta\tau}(2-\beta\tau)) - (1-e^{\beta\tau}+\beta\tau)v_t)\rho\sigma \\
K^{1,1,1,1} &= 3\sigma^2(\alpha-2v_t)\frac{(1-e^{\beta\tau})^2}{2\beta^3 e^{2j\beta\tau}} + 6\sigma^2(\alpha-v_t)\frac{\tau^2\rho^2\beta^2-(2\rho^2+1)(e^{\beta\tau}-\beta\tau-1)}{\beta^3 e^{j\beta\tau}} \\
&\quad + 3\alpha\sigma^2\frac{1+e^{\beta\tau}(\beta\tau-1)-8\rho^2(e^{\beta\tau}-1)+4\rho^2\tau\beta(1+e^{\beta\tau})}{\beta^3 e^{\beta\tau}} \\
K^{1,2} &= \frac{e^{-\beta\tau}}{\beta}(\alpha(e^{\beta\tau}-1-\beta\tau)+\beta\tau v_t)\rho\sigma \\
K^{1,1,2} &= \frac{e^{-2\beta\tau}}{2\beta^2}(2(1-e^{\beta\tau}(1-\beta\tau-\beta^2\tau^2\rho^2))v_t - \alpha(1-e^{2\beta\tau}(1+4\rho^2) \\
&\quad + 2e^{\beta\tau}(\beta\tau+2\rho^2+2\beta\tau\rho^2+\beta^2\tau^2\rho^2)))\sigma^2 \\
K^{1,2,2} &= \frac{e^{-2\beta\tau}}{\beta^2}(\alpha(e^{\beta\tau}-1)(e^{\beta\tau}-1-\beta\tau)-(1+2\beta\tau-e^{\beta\tau}(1+\beta\tau))v_t)\rho\sigma^3 \\
K^{1,1,2,2} &= \frac{e^{-3\beta\tau}}{2\beta^3}((-3+4e^{\beta\tau}(1-\beta\tau-\rho^2-2\beta\tau\rho^2-2\beta^2\tau^2\rho^2)-e^{2\beta\tau}(1-2\beta\tau-4\rho^2 \\
&\quad - 4\beta\tau\rho^2-2\beta^2\tau^2\rho^2))v_t + \alpha(1+e^{3\beta\tau}(1+6\rho^2)-e^{2\beta\tau}(1+12\rho^2 \\
&\quad + 2\beta^2\tau^2\rho^2+2\beta(\tau+4\tau\rho^2))+e^{\beta\tau}(-1+6\rho^2+4\beta^2\tau^2\rho^2+2\beta(\tau+4\tau\rho^2))))\sigma^4.
\end{aligned}$$

The cumulants with respect to v_t with various orders are given in Example 3.2 for the square-root diffusion process. Substituting the cumulants into (2.11) for the multivariate case, it is straightforward to derive the necessary estimating equations.

4. MONTE CARLO SIMULATIONS

In this section, we investigate via simulation the performance of the estimation method proposed in the paper for the Markov processes in comparison with alternative estimation methods. In order to focus on the relative performance of various estimators, we restrict our simulations to the models with exact sampling paths. As a result we focus on the univariate continuous-time square-root diffusion process and the bivariate discrete-time stochastic volatility process. Through these two models, we demonstrate the performance of our estimation method for both univariate and bivariate Markov processes and in both discrete time and continuous time frameworks.

4.1. The continuous-time square-root diffusion process

The square-root diffusion process is specified in (3.2). The parameter values are set as $\alpha = 0.075$, $\beta = 0.80$, $\sigma = 0.100$, which are close to the estimates of an interest rate process using historical U.S. three-month Treasury bill yields. The choice of parameter values gives an integer value of the degree of freedom for the non-central chi-square transition density function and makes it feasible to generate exact sampling paths.[2] Thus, there is no discretization error and differences between different estimates are entirely due to the different estimation methods. We set two sampling intervals, i.e. $\Delta = 1/4, 1$ with sample size $T = 250, 500$. In each sampling path simulation, the first 200 observations are discarded to mitigate the start-up effect. The number of replications in the Monte Carlo simulation is 1000. The estimation methods we consider include the C-AML developed in this paper, the MLE, QMLE and GMM.

MLE Solving from the Kolmogorov backward (or Fokker–Planck) equation or from the CCF via Fourier inversion, the transition density function of the square-root process can be obtained as

$$f(x_t \mid x_{t-\tau}; \theta) = ce^{-u-v} \left(\frac{v}{u}\right)^{q/2} I_q(2(uv)^{1/2}), \tag{4.1}$$

with x_t taking non-negative values, where $c = 2\beta/(\sigma^2(1 - e^{-\beta\tau}))$, $u = cx_{t-\tau}e^{-\beta t}$, $v = cx_t$, $q = \frac{2\beta\alpha}{\sigma^2} - 1$, 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, $\chi^2[2cx_t; 2q + 2, 2u]$, with $2q + 2$ degrees of freedom and parameter of non-centrality $2u$ proportional to the current level of the stochastic process. 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. $g(x_t; \theta) = \frac{\omega^s}{\Gamma(s)} x_t^{s-1} e^{-\omega x_t}$, where $\omega = 2\beta/\sigma^2$ and $s = 2\alpha\beta/\sigma^2$, with mean α and variance $\alpha\sigma^2/2\beta$. The MLE is based on

$$\frac{\partial \ln g(x_1; \theta)}{\partial \theta} + \sum_{t=1}^{T} \frac{\partial \ln f(x_{t+1} \mid x_t; \theta)}{\partial \theta} = 0.$$

[2] The square-root diffusion process specified in (3.2) has a non-central chi-square transition density function with a degree of freedom $2q + 2$, where $q = \frac{2\beta\alpha}{\sigma^2} - 1$. Given the parameter values used in our simulation, q equals 11.

QMLE The QMLE is based on the conditional mean and variance which are given in Example 3.2 of Section 3 as well as the above unconditional mean and variance of the square-root process. As mentioned earlier, QMLE is equivalent to the C-AMLE proposed in this paper when the Edgeworth/Gram–Charlier expansion has an order of 2, namely $p = 2$.

GMM The same moment conditions as in Chan et al. (1992) are used for GMM, except that the moment conditions are derived from the continuous-time model and are not subject to discretization bias, i.e.

$$f_t(\theta) = \begin{bmatrix} \epsilon_t \\ \epsilon_t^2 - E\left[\epsilon_t^2 \,\middle|\, x_{t-\Delta}\right] \end{bmatrix}, \tag{4.2}$$

where $t = 1, 2, \ldots, T, \theta = (\alpha, \beta, \sigma)$ and the lagged variable is used as instrumental variable in the estimation, where $\epsilon_t = \Delta x_t - E[\Delta x_t \,|\, x_{t-\Delta}]$ with $E[\Delta x_t \,|\, x_{t-\Delta}] = (1 - e^{-\beta\Delta})(\alpha - x_{t-\Delta})$, $E[\epsilon_t^2 \,|\, x_{t-\Delta}] = \frac{\sigma^2}{\beta}(e^{-\beta\Delta} - e^{-2\beta\Delta})x_{t-\Delta} + \frac{\alpha\sigma^2}{2\beta}(1 - e^{-\beta\Delta})^2$.

The simulation results of alternative estimators are reported in Table 1 for different sampling intervals and sample sizes. Certain interesting observations for the estimation of the square-root process are worth noting. The unconditional mean parameter (α) tends to be straightforward to estimate. Measured by both bias and mean squared error, all four methods have similar performance. The estimation of the conditional variance parameter (σ) improves as both sample size and sampling frequency increases. Among alternative estimators, QMLE has the worst performance. The mean-reverting parameter (β) turns out to be most difficult to estimate. Overall, there tends to be an upward bias for all estimators. As the sampling period increases, the upward bias reduces. In other words, the estimation of β requires a large sample to be accurate.

Overall, the C-AML estimator performs as well as the ML estimator and consistently outperforms other estimators. The difference with the QML and GMM suggests that the moments with order higher than 3 are informative and help to improve the parameter estimation in our estimation procedure. On the other hand, we also perform GMM estimation using the same moment conditions as those used in the C-AML procedure. The performance of the GMM estimation generally deteriorates. For example, with sampling interval $\Delta = 1/4$ and sample size $N = 250$, the root mean square error of β estimates increases to 0.2302 (from 0.2211 as reported in Table 1). The deterioration is mainly caused by the increase in the standard error of the parameter estimates. Finally, it is noted that since MLE involves numerical evaluation of the Bessel function of the first kind, computationally the alternative estimators, including C-AML estimator, all have a certain advantage.

4.2. The discrete-time stochastic volatility process

The discrete-time stochastic volatility process is specified in (3.5). We consider two sets of parameter values, namely $\{\alpha, \beta, \sigma, \rho\} = \{-0.35, 0.95, 0.25, -0.50; -0.70, 0.90, 0.35, -0.25\}$ which are similar to the values used in Andersen and Sørensen (1996) except that we set non-zero values for ρ in our study. Again, there is no approximation error involved in the path simulation and differences between different estimates are entirely due to the different estimation methods. We set two sample sizes $T = 250, 500$. In each sampling path simulation, the first 200 observations are discarded to mitigate the start-up effect. The number of replications in the Monte Carlo simulation is 1000. Since neither the transition density function nor the CCF is available in closed form for the SV process, we only include the QMLE and GMM in our comparison.

Table 1. Monte Carlo simulation results of the square-root diffusion model.

Parameter	Estimation method	Mean	Median	SD	$\sqrt{m.s.e.}$	(95 percentiles)
Panel A: Sampling Interval $\Delta = 1/4$, Sample Size = 250						
α	C-AML	0.0748	0.0746	0.0043	0.0043	(0.0671 0.0829)
(0.075)	ML	0.0748	0.0746	0.0043	0.0043	(0.0672 0.0840)
	QML	0.0748	0.0746	0.0043	0.0043	(0.0672 0.0838)
	GMM	0.0747	0.0746	0.0044	0.0044	(0.0667 0.0839)
β	C-AML	0.8925	0.8797	0.2002	0.2205	(0.5535 1.3618)
(0.800)	ML	0.8958	0.8803	0.1991	0.2208	(0.5486 1.3545)
	QML	0.8929	0.8790	0.2089	0.2285	(0.5221 1.3691)
	GMM	0.8840	0.8692	0.2046	0.2211	(0.5465 1.3476)
σ	C-AML	0.0995	0.0994	0.0050	0.0051	(0.0894 0.1084)
(0.100)	ML	0.0994	0.0994	0.0051	0.0051	(0.0896 0.1100)
	QML	0.0712	0.0712	0.0038	0.0291	(0.0642 0.0792)
	GMM	0.0997	0.0996	0.0055	0.0055	(0.0894 0.1103)
Panel B: Sampling Interval $\Delta = 1/4$, Sample Size = 500						
α	C-AML	0.0748	0.0747	0.0031	0.0031	(0.0686 0.0802)
(0.075)	ML	0.0748	0.0746	0.0031	0.0031	(0.0687 0.0811)
	QML	0.0748	0.0746	0.0031	0.0031	(0.0688 0.0811)
	GMM	0.0747	0.0746	0.0031	0.0031	(0.0687 0.0810)
β	C-AML	0.8508	0.8420	0.1374	0.1465	(0.6133 1.1521)
(0.800)	ML	0.8517	0.8427	0.1370	0.1464	(0.6094 1.1480)
	QML	0.8453	0.8337	0.1450	0.1519	(0.5921 1.1606)
	GMM	0.8428	0.8301	0.1403	0.1466	(0.6016 1.1400)
σ	C-AML	0.0995	0.0991	0.0034	0.0035	(0.0922 0.1050)
(0.100)	ML	0.0990	0.0990	0.0035	0.0036	(0.0924 0.1056)
	QML	0.0709	0.0709	0.0026	0.0292	(0.0659 0.0759)
	GMM	0.0998	0.0997	0.0037	0.0037	(0.0927 0.1074)
Panel C: Sampling Interval $\Delta = 1$, Sample Size = 250						
α	C-AML	0.0749	0.0749	0.0023	0.0023	(0.0705 0.0793)
(0.075)	ML	0.0749	0.0749	0.0022	0.0022	(0.0709 0.0796)
	QML	0.0749	0.0749	0.0022	0.0022	(0.0709 0.0796)
	GMM	0.0750	0.0749	0.0023	0.0023	(0.0708 0.0797)
β	C-AML	0.8333	0.8236	0.1424	0.1462	(0.5896 1.1490)
(0.800)	ML	0.8385	0.8246	0.1364	0.1417	(0.5955 1.1497)
	QML	0.8376	0.8203	0.1473	0.1520	(0.5883 1.1710)
	GMM	0.8327	0.8200	0.1438	0.1474	(0.5861 1.1631)

Table 1. Continued.

Parameter	Estimation method	Mean	Median	SD	$\sqrt{m.s.e.}$	(95 percentiles)
σ	C-AML	0.1003	0.0997	0.0067	0.0067	(0.0881 0.1137)
(0.100)	ML	0.1002	0.1000	0.0066	0.0066	(0.0883 0.1138)
	QML	0.0714	0.0711	0.0050	0.0291	(0.0625 0.0818)
	GMM	0.1001	0.0997	0.0070	0.0070	(0.0874 0.1143)
Panel D: Sampling Interval $\Delta = 1$, Sample Size = 500						
α	C-AML	0.0750	0.0749	0.0015	0.0015	(0.0718 0.0779)
(0.075)	ML	0.0749	0.0748	0.0016	0.0016	(0.0718 0.0781)
	QML	0.0749	0.0748	0.0016	0.0016	(0.0718 0.0781)
	GMM	0.0749	0.0749	0.0016	0.0016	(0.0718 0.0781)
β	C-AML	0.8145	0.8131	0.0991	0.1002	(0.6452 1.0244)
(0.800)	ML	0.8212	0.8146	0.0977	0.0999	(0.6531 1.0294)
	QML	0.8173	0.8113	0.1051	0.1065	(0.6329 1.0490)
	GMM	0.8151	0.8108	0.0998	0.1009	(0.6451 1.0286)
σ	C-AML	0.0998	0.0996	0.0047	0.0047	(0.0907 0.1090)
(0.100)	ML	0.0997	0.0996	0.0047	0.0047	(0.0911 0.1093)
	QML	0.0710	0.0708	0.0036	0.0293	(0.0644 0.0784)
	GMM	0.0999	0.0997	0.0049	0.0049	(0.0906 0.1098)

QMLE The QMLE is based on the conditional mean and variance of x_t and h_t as well as their cross moment which are given in Example 3.5 of Section 3 as well as the unconditional mean and variance of x_t and h_t. Again, QMLE is equivalent to the C-AMLE proposed in this paper when the Edgeworth/Gram–Charlier expansion has a order of 2, namely $p = 2$.

GMM The following moment conditions are used in the GMM estimation:

$$f_t(\theta) = \begin{bmatrix} \nu_t \\ \nu_t^2 - E\left[\nu_t^2 \,\middle|\, h_{t-1}\right] \\ \nu_t x_t - E\left[\nu_t x_t \,\middle|\, h_{t-1}\right] \end{bmatrix}, \tag{4.3}$$

where $\nu_t = (h_t - \alpha - \beta h_{t-1})/\sigma, t = 1, 2, \ldots, T, \theta = (\alpha, \beta, \sigma)$ and the lagged variable h_{t-1} is used as instrumental variable in the estimation.

The simulation results for alternative estimators are reported in Table 2 for different parameter sets and sample sizes. The most noticeable difference between C-AML and the QMLE and GMM estimators is the performance of the ρ parameter estimation. As measured by both bias and root mean squared error, the C-AML estimator outperforms both QMLE and GMM, suggesting that higher moment conditions help to identify the correlation parameter and improve its estimation. It is not surprising that the estimation of α, β and σ has similar performance for all three estimators as h_t follows a simple AR(1) Gaussian process. We note that in our simulations, the simulated volatility process is also used in estimation. This is different from Andersen and Sørensen (1996) where volatility is a latent variable. For this reason, the non-convergence issue encountered in our simulations is much less severe than in Andersen and Sørensen (1996) and the relative performance of alternative estimators is not affected even when we include the non-convergence cases. Our further analysis of the GMM estimation also shows

Table 2. Monte Carlo simulation results of the discrete-time SV model.

Parameter	Estimation method	Mean	Median	SD	$\sqrt{m.s.e.}$	(95 percentiles)
Panel A: Sample Size = 250, Parameter Set I						
α	C-AML	−0.4321	−0.4207	0.1110	0.1380	(−0.6701 −0.2543)
(−0.35)	QML	−0.4626	−0.4501	0.1322	0.1736	(−0.7529 −0.2533)
	GMM	−0.4573	−0.4435	0.1318	0.1699	(−0.7427 −0.2497)
β	C-AML	0.9383	0.9398	0.0158	0.0197	(0.9045 0.9637)
(0.95)	QML	0.9339	0.9362	0.0189	0.0248	(0.8926 0.9638)
	GMM	0.9347	0.9367	0.0188	0.0243	(0.8936 0.9642)
σ	C-AML	0.2481	0.2484	0.0039	0.0044	(0.2399 0.2550)
(0.25)	QML	0.2476	0.2481	0.0043	0.0050	(0.2384 0.2548)
	GMM	0.2473	0.2478	0.0044	0.0051	(0.2379 0.2543)
ρ	C-AML	−0.4779	−0.4804	0.1297	0.1316	(−0.7136 −0.2192)
(−0.50)	QML	−0.4459	−0.4397	0.1385	0.1486	(−0.7267 −0.2044)
	GMM	−0.4460	−0.4384	0.1398	0.1498	(−0.7332 −0.2095)
Panel B: Sample Size = 500, Parameter Set I						
α	C-AML	−0.3978	−0.3864	0.0810	0.0940	(−0.5665 −0.2692)
(−0.35)	QML	−0.4119	−0.3982	0.0914	0.1104	(−0.6171 −0.2665)
	GMM	−0.4086	−0.3953	0.0913	0.1084	(−0.6123 −0.2636)
β	C-AML	0.9432	0.9447	0.0116	0.0134	(0.9185 0.9617)
(0.95)	QML	0.9412	0.9432	0.0131	0.0158	(0.9118 0.9621)
	GMM	0.9417	0.9437	0.0131	0.0155	(0.9126 0.9623)
σ	C-AML	0.2491	0.2493	0.0021	0.0022	(0.2449 0.2526)
(0.25)	QML	0.2489	0.2490	0.0023	0.0026	(0.2442 0.2528)
	GMM	0.2487	0.2488	0.0023	0.0026	(0.2440 0.2524)
ρ	C-AML	−0.5023	−0.5031	0.1014	0.1014	(−0.6512 −0.2918)
(−0.50)	QML	−0.4727	−0.4697	0.1098	0.1131	(−0.6965 −0.2762)
	GMM	−0.4726	−0.4700	0.1096	0.1129	(−0.6934 −0.2797)
Panel C: Sample Size = 250, Parameter Set II						
α	C-AML	−0.7970	−0.7781	0.1720	0.1975	(−1.1476 −0.5183)
(−0.70)	QML	−0.8093	−0.7955	0.1705	0.2024	(−1.1717 −0.5315)
	GMM	−0.7998	−0.7863	0.1706	0.1975	(−1.1569 −0.5161)
β	C-AML	0.8861	0.8888	0.0245	0.0282	(0.8364 0.9258)
(0.90)	QML	0.8845	0.8867	0.0243	0.0288	(0.8338 0.9242)
	GMM	0.8857	0.8877	0.0243	0.0282	(0.8348 0.9260)
σ	C-AML	0.3476	0.3480	0.0055	0.0060	(0.3370 0.3572)
(0.35)	QML	0.3465	0.3471	0.0058	0.0068	(0.3352 0.3565)
	GMM	0.3467	0.3471	0.0057	0.0066	(0.3355 0.3566)

Table 2. Continued.

Parameter	Estimation method	Mean	Median	SD	$\sqrt{m.s.e.}$	(95 percentiles)
ρ	C-AML	−0.2473	−0.2354	0.1082	0.1082	(−0.4202 −0.0229)
(−0.25)	QML	−0.2276	−0.2308	0.1120	0.1142	(−0.4371 −0.0143)
	GMM	−0.2289	−0.2304	0.1137	0.1156	(−0.4415 −0.0213)
Panel D: Sample Size = 500, Parameter Set II						
α	C-AML	−0.7430	−0.7401	0.1148	0.1225	(−0.9844 −0.5512)
(−0.70)	QML	−0.7491	−0.7415	0.1189	0.1286	(−0.9889 −0.5493)
	GMM	−0.7447	−0.7344	0.1186	0.1267	(−0.9877 −0.5456)
β	C-AML	0.8939	0.8944	0.0164	0.0175	(0.8595 0.9218)
(0.90)	QML	0.8930	0.8943	0.0170	0.0183	(0.8590 0.9212)
	GMM	0.8936	0.8950	0.0169	0.0181	(0.8593 0.9219)
σ	C-AML	0.3485	0.3487	0.0030	0.0034	(0.3424 0.3537)
(0.35)	QML	0.3479	0.3481	0.0033	0.0039	(0.3413 0.3535)
	GMM	0.3479	0.3482	0.0033	0.0039	(0.3414 0.3534)
ρ	C-AML	−0.2463	−0.2437	0.0779	0.0780	(−0.3710 −0.0925)
(−0.25)	QML	−0.2365	−0.2331	0.0845	0.0855	(−0.3990 −0.0872)
	GMM	−0.2368	−0.2337	0.0843	0.0853	(−0.4002 −0.0869)

that adding higher-order moments such as $E[x_t^3 \mid h_{t-1}]$ or $E[x_t v_t^2 \mid h_{t-1}]$ to those in (4.3) helps to identify the parameter ρ. However, by adding these extra high-order moment conditions in the GMM estimation, there is in general an increase in the standard errors of the parameter estimates. As noted in Andersen and Sørensen (1996), this is likely due to the deterioration in the estimation of the weighting matrix as the number of moments increases. We also note that as shown theoretically by Newey and Smith (2004), increasing the number of moment conditions will increase the bias but decrease the standard error of the GMM estimator. To reconcile these two observations, we follow Andersen and Sørensen (1996) and perform the following simulations. First, we simulate 10,000 sample observations and estimate the so-called 'true' optimal weighting matrix in the GMM procedure. This 'true' optimal weighting matrix is then used in our simulation with different sampling frequencies and sample sizes. In all four cases of Table 2, we indeed observe a decrease in the standard errors of the GMM estimates. The results confirm the following fundamental trade-off documented in Andersen and Sørensen (1996): inclusion of additional moments improves estimation performance for a given degree of precision in the estimation of the weighting matrix, but in finite samples this must be balanced against the deterioration in the estimate of the weighting matrix as the number of moments increases. Since the C-AMLE procedure has a specific analytical structure for moment restrictions as given in the estimating equations, the weighting matrix is analytically available and there is no need to estimate the weighting matrix numerically. This further underlines the advantage of our estimator over the GMM and explains why our estimator achieves better performance with finite samples.

5. CONCLUSION

In this paper, we develop a new estimator for Markov models by combining an approximation to the logarithmic transition density along with the first-order conditions associated with the ECF estimation approach. The estimator is consistent. As an alternative to the MLE due to the unavailability of the exact likelihood function, the new approach is exact and parsimonious. The method applies to processes with closed-form CCF, e.g. the common continuous-time affine diffusion and jump-diffusion processes, the discrete time CAR processes introduced by Darolles et al. (2006). Furthermore, since only the conditional cumulants are required in the estimation the method also applies to processes which have analytical expressions for certain orders of conditional cumulants. Using various examples, we illustrate the application of our approach to Markov processes in both discrete time and continuous time and for both univariate and multivariate cases. The Monte Carlo simulations for selected models confirm that the estimator has desirable finite sample performance in comparison with other methods.

ACKNOWLEDGMENTS

We wish to thank Yacine Aït-Sahalia, Alan Rogers, Peter Thompson, Stéphane Gregoir (the Editor), and two anonymous referees for helpful comments and suggestions, along with participants at the 2003 North American Econometric Society Meetings in Atlanta, University of Toronto Economics Workshop, and the 2004 Financial Econometrics Conference at the University of Waterloo. Both authors acknowledge financial support from NSERC, Canada. The usual disclaimer applies.

REFERENCES

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed-form approach. *Econometrica 70*, 223–62.

Aït-Sahalia, Y. (2003). Closed-form likelihood expansions for multivariate diffusions. Working paper, Princeton University and NBER.

Andersen, T. G., L. Benzoni and J. Lund (2002). An empirical investigation of continuous-time equity return models. *Journal of Finance 57*, 1239–84.

Andersen, T. G., T. Bollerslev, F. X. Diebold and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association 96*, 42–55.

Andersen, T. G. and B. E. Sørensen (1996). GMM estimation of a stochastic volatility model: a Monte Carlo study. *Journal of Business and Economic Statistics 14*, 328–52.

Barndorff-Nielsen, O. E. and N. Shephard (2002a). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society*, Series B *64*, 252–80.

Barndorff-Nielsen, O. E. and N. Shephard (2002b). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics 17*, 457–77.

Bollerslev, T. and J. Wooldridge (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews 11*, 143–72.

Carrasco, M., M. Chernov, J.-P. Florens and E. Ghysels (2007). Efficient estimation of jump diffusions and general dynamic models with a continuum of moment conditions. *Journal of Econometrics 140*, 529–73.

Carrasco, M. and J. Florens (2000). Efficient GMM estimation using the empirical characteristic function. Document du travail 2000–33, CREST, Paris.

Chacko, G. and L. M. Viceira (2003). Spectral GMM estimation of continuous-time processes. *Journal of Econometrics 116*, 259–92.

Chan, K., G. A. Karolyi, F. A. Longstaff and A. B. Sanders (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance 47*, 1209–27.

Chernov, M. and E. Ghysels (2000). A study toward a unified approach to the joint estimation of objective and risk-neutral measures for the purposes of options valuation. *Journal of Financial Economics 56*, 407–58.

Cramér, H. (1925). On some classes of series used in mathematical statistics. *Proceedings of the Sixth Scandinavian Congress of Mathematicians*, Copenhagen, 399–425.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton: Princeton University Press.

Darolles, S., C. Gouriéroux and J. Jasiak (2006). Structural Laplace transform and compound autoregressive models. *Journal of Time Series Analysis 27*, 477–503.

Duffie, D. and K. J. Singleton (1993). Simulated moments estimation of Markov models of asset prices. *Econometrica 61*, 929–52.

Feuerverger, A. (1990). An efficiency result for the empirical characteristic function in stationary time-series models. *Canadian Journal of Statistics 18*, 155–61.

Feuerverger, A. and P. McDunnough (1981). On some Fourier methods for inference. *Journal of the American Statistical Association 76*, 379–87.

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

Fisher, M. and C. Gilles (1996). Estimating exponential affine models of the term structure. Working paper, Federal Reserve Atlanta.

Gallant, A. and J. Long (1997). Estimating stochastic differential equations efficiently by minimum chi-square. *Biometrika 84*, 125–41.

Gallant, A. R. and G. Tauchen (1996). Which moments to match? *Econometric Theory 12*, 657–81.

Gouriéroux, C. and J. Jasiak (2000). Compound gamma processes. Working paper, CREST, Paris.

Gouriéroux, C., A. Monfort and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics 8*, S85–S199.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimator. *Econometrica 50*, 1029–54.

Hansen, L. P. and J. Scheinkman (1996). Back to the future: generating moment implications for continuous-time Markov processes. *Econometrica 63*, 767–804.

Heston, S. L. (1993). A closed form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies 6*, 327–44.

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

Jones, C. S. (1998). Bayesian estimation of continuous-time finance models. Working paper, University of Rochester.

Kendall, M. and A. Stuart (1977). *The Advanced Theory of Statistics,* Volume 1. London: Charles Griffin.

Knight, J. and S. E. Satchell (1997). The cumulant generating function method estimation, implementation and asymptotic efficiency. *Econometric Theory 13*, 170–84.

Liu, J. (1997). Generalized method of moments estimation of affine diffusion processes. Working Paper, Graduate School of Business, Stanford University.

McCullagh, P. (1987). *Tensor Methods in Statistics*. London: Chapman and Hall.

Newey, W. and R. Smith (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica 72*, 219–55.

Schmidt, P. (1982). An improved version of the Quandt–Ramsey MGF estimator for mixtures of normal distributions and switching regressions. *Econometrica 6*, 501–24.

Singleton, K. J. (2001). Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics 102*, 111–41.

Skovgaard, M. (1986). On multivariate Edgeworth expansions. *International Statistical Review 54*, 169–86.

Yu, J. (1998). Empirical characteristic function in time series estimation and a test statistic in financial modelling. Unpublished Ph.D dissertation, University of Western Ontario.

APPENDIX A: PROOFS OF LEMMAS AND THEOREM

Proof of Lemma 2.1: The proof is essentially along the lines of Singleton (2001). From (2.2), we have

$$\frac{\partial \ln f(x_{t+1} \mid x_t; \theta)}{\partial \theta} = \int e^{ir' x_{t+1}} w(r, t \mid x_t; \theta)\, dr$$

and

$$\begin{aligned}
&\int w(r, t \mid x_t; \theta) \phi(r, x_{t+1} \mid x_t; \theta)\, dr \\
&= \int w(r, t \mid x_t; \theta) \int e^{ir' x_{t+1}} f(x_{t+1} \mid x_t; \theta)\, dx_{t+1}\, dr \\
&= \int \left(\int w(r, t \mid x_t; \theta) e^{ir' x_{t+1}} dr \right) f(x_{t+1} \mid x_t; \theta)\, dx_{t+1} \\
&= E\left[\int e^{ir' x_{t+1}} w(r, t \mid x_t; \theta)\, dr \right] \\
&= E\left[\frac{\partial \ln f(x_{t+1} \mid x_t; \theta)}{\partial \theta} \right]
\end{aligned}$$

Thus the estimating equations (2.1) lead to

$$\frac{1}{T} \sum_{t=1}^{T} \left[\frac{\partial \ln f(x_{t+1} \mid x_t; \theta)}{\partial \theta} - E\left[\frac{\partial \ln f(x_{t+1} \mid x_t; \theta)}{\partial \theta} \,\middle|\, x_t \right] \right] = 0,$$

which is equivalent as ML estimation. □

Proof of Theorem 2.1: Similar to the manipulation to equation (2.8) as in the proof of Lemma 2.1 except $w(r, t \mid x_t; \theta)$ being replaced with $\hat{w}(r, t \mid x_t; \theta)$, we have equation (2.9). Firstly, from (2.8), we note immediately from Singleton (2001) that under regularity conditions our estimator is consistent and asymptotically normally distributed. Secondly, denote equation (2.8) as

$$H(\theta) = \frac{1}{T} \sum_{t=1}^{T} h_t(\theta) = 0$$

under certain regularity conditions, for a fixed p (the order in Edgeworth/Gram–Charlier expansion) we have that the asymptotic variance–covariance matrix Ω_p is given by

$$\Omega_p = D(\theta)^{-1} \Sigma(\theta) D(\theta)^{-1},$$

with $D(\theta) = \text{plim} \frac{1}{T} \sum_{t=1}^{T} \frac{\partial h_t(\theta)}{\partial \theta}$ and $\Sigma(\theta) = \text{plim} T H(\theta) H(\theta)'$. Refer to equation (2.9), we have

$$D(\theta) = \text{plim} \frac{1}{T} \sum_{t=1}^{T} \left[\frac{\partial^2 \ln \hat{f}_p(x_{t+1} \mid x_t; \theta)}{\partial \theta \partial \theta'} - \frac{\partial}{\partial \theta'} E\left[\frac{\partial \ln \hat{f}_p(x_{t+1} \mid x_t; \theta)}{\partial \theta} \,\middle|\, x_t \right] \right].$$

Since

$$E\left[\frac{\partial \ln \hat{f}_p(x_{t+1}\,|\,x_t;\theta)}{\partial\theta}\,\middle|\,x_t\right] = \int \frac{\partial \ln \hat{f}_p(x_{t+1}\,|\,x_t;\theta)}{\partial\theta}\cdot f(x_{t+1}\,|\,x_t;\theta)\,dx_{t+1},$$

thus

$$\begin{aligned}\frac{\partial}{\partial\theta'}E\left[\frac{\partial \ln \hat{f}_p(x_{t+1}\,|\,x_t;\theta)}{\partial\theta}\,\middle|\,x_t\right] &= \int \frac{\partial^2 \ln \hat{f}_p}{\partial\theta\partial\theta'}\cdot f(x_{t+1}\,|\,x_t;\theta)\cdot dx_{t+1} + \int \frac{\partial \ln \hat{f}_p}{\partial\theta}\cdot\frac{\partial f}{\partial\theta}\cdot dx_{t+1}\\ &= E\left[\frac{\partial^2 \ln \hat{f}_p}{\partial\theta\partial\theta'}\,\middle|\,x_t\right] + \int \frac{\partial \ln \hat{f}_p}{\partial\theta}\cdot\frac{\partial \ln f}{\partial\theta}\cdot f\cdot dx_{t+1}\\ &= E\left[\frac{\partial^2 \ln \hat{f}_p}{\partial\theta\partial\theta'}\,\middle|\,x_t\right] + E\left[\frac{\partial \ln \hat{f}_p}{\partial\theta}\cdot\frac{\partial \ln f}{\partial\theta}\,\middle|\,x_t\right].\end{aligned}$$

Furthermore

$$\begin{aligned}E\left[\frac{\partial^2 \ln \hat{f}_p}{\partial\theta\partial\theta'}\,\middle|\,x_t\right] &= E\left[\frac{\partial}{\partial\theta'}\left(\frac{1}{\hat{f}_p}\cdot\frac{\partial \hat{f}_p}{\partial\theta}\right)\,\middle|\,x_t\right]\\ &= E\left[-\frac{1}{\hat{f}_p^2}\cdot\frac{\partial \hat{f}_p}{\partial\theta}\cdot\frac{\partial \hat{f}_p}{\partial\theta'} + \frac{1}{\hat{f}_p}\cdot\frac{\partial^2 \hat{f}_p}{\partial\theta\partial\theta'}\,\middle|\,x_t\right]\\ &= -E\left[\frac{\partial \ln \hat{f}_p}{\partial\theta}\cdot\frac{\partial \ln \hat{f}_p}{\partial\theta'}\,\middle|\,x_t\right] + \int \frac{1}{\hat{f}_p}\cdot\frac{\partial^2 \hat{f}_p}{\partial\theta\partial\theta'}\cdot f\cdot dx_{t+1}\\ &\neq -E\left[\frac{\partial \ln \hat{f}_p}{\partial\theta}\cdot\frac{\partial \ln \hat{f}_p}{\partial\theta'}\,\middle|\,x_t\right].\end{aligned}$$

Note, the last inequality holds if $\ln \hat{f}_p \neq \ln f$. Consequently,

$$\begin{aligned}D(\theta) = \text{plim}\frac{1}{T}\sum_{t=1}^{T}\Bigg\{&\frac{\partial^2 \ln \hat{f}_p(x_{t+1}\,|\,x_t;\theta)}{\partial\theta\partial\theta'} - E\left[\frac{\partial \ln \hat{f}_p}{\partial\theta}\cdot\left(\frac{\partial \ln f}{\partial\theta'} - \frac{\partial \ln \hat{f}_p}{\partial\theta'}\right)\,\middle|\,x_t\right]\\ &-\int \frac{1}{\hat{f}_p}\cdot\frac{\partial^2 \hat{f}_p}{\partial\theta\partial\theta'}\cdot f\cdot dx_{t+1}\Bigg\}.\end{aligned}$$

On the other hand,

$$\begin{aligned}\Sigma(\theta) &= \text{plim}\frac{1}{T}\sum\sum h(\theta)h(\theta)'\\ &= \text{plim}\frac{1}{T}\sum\Bigg\{\frac{\partial \ln \hat{f}_p}{\partial\theta}\frac{\partial \ln \hat{f}_p}{\partial\theta'} - E\left[\frac{\partial \ln \hat{f}_p}{\partial\theta}\,\middle|\,x_t\right]\frac{\partial \ln \hat{f}_p}{\partial\theta'}\\ &\quad -\frac{\partial \ln \hat{f}_p}{\partial\theta}E\left[\frac{\partial \ln \hat{f}_p}{\partial\theta'}\,\middle|\,x_t\right] + E\left[\frac{\partial \ln \hat{f}_p}{\partial\theta}\,\middle|\,x_t\right]E\left[\frac{\partial \ln \hat{f}_p}{\partial\theta'}\,\middle|\,x_t\right]\Bigg\}.\end{aligned}$$

Note that the above results are derived for a fixed p which is the truncation order of the Gram–Charlier series expansion. If as p goes to infinity we have $\ln \hat{f}_p \to \ln f$, i.e. the expansion is convergent, then we have

$$D(\theta) = \text{plim}\frac{1}{T}\sum\frac{\partial^2 \ln f(x_{t+1}\,|\,x_t;\theta)}{\partial\theta\partial\theta'}$$

and

$$\Sigma(\theta) = \text{plim} \frac{1}{T} \sum \frac{\partial \ln f(x_{t+1} \mid x_t; \theta)}{\partial \theta} \frac{\partial \ln f(x_{t+1} \mid x_t; \theta)}{\partial \theta'}.$$

In other words, $D(\theta) = \Sigma(\theta) = I(\theta)$ and the asymptotic covariance is equal to $I(\theta)^{-1}$. In this case, the estimator becomes the ECF estimator proposed in Singleton (2001) with optimal weight function and thus achieves ML efficiency. □

Proof of Lemma 2.2: The results in Lemma 2.2 follow immediately from the substitution of equation (2.6) into equation (2.9). Alternatively, the estimating equation (2.11) can be derived by substituting the approximating weight function into equation (2.8) and apply the definition of cumulants. □

APPENDIX B: CCF OF BIVARIATE OU PROCESS AND SV PROCESS

CCF of Bivariate OU Process. The joint characteristic function of $(y_{t+\tau}, x_{t+\tau})$ conditional on $\mathcal{F}_t$ can be written as

$$\begin{aligned} \psi(r_1, r_2; y_{t+\tau}, x_{t+\tau} \mid y_t, x_t) &= E[\exp\{ir_1 y_{t+\tau} + ir_2 x_{t+\tau}\} \mid y_t, x_t] \\ &= \exp\{C(\tau; r_1, r_2) + D1(\tau; r_1, r_2)' \, y_t + D2(\tau; r_1, r_2)' \, x_t\}, \end{aligned}$$

where $C(\cdot)$, $D1(\cdot)$ and $D2(\cdot)$ are solved from the Ricatti equations as

$$\begin{aligned} C(\tau; r_1, r_2) &= \frac{1}{4\kappa} r_1^2 \left(\sigma^2 + \frac{\kappa^2}{(\beta - \kappa)^2} \sigma_x^2 \right) (e^{-2\kappa\tau} - 1) + \frac{1}{4\beta} \left(r_2 - r_1 \frac{\kappa}{\beta - \kappa} \right) \sigma_x^2 (e^{-2\beta\tau} - 1) \\ &\quad + r_1 \left(r_2 - r_1 \frac{\kappa}{\beta - \kappa} \right) \frac{\kappa}{\beta^2 - \kappa^2} \sigma_x^2 (e^{-(\beta+\kappa)\tau} - 1) \\ D1(\tau; r_1, r_2) &= ir_1 e^{-\kappa\tau} \\ D2(\tau; r_1, r_2) &= i \left(r_2 - r_1 \frac{\kappa}{\beta - \kappa} (e^{(\beta-\kappa)\tau} - 1) \right) e^{-\beta\tau}. \end{aligned}$$

CCF of Square-Root SV Process. The joint characteristic function of $(x_{t+\tau}, v_{t+\tau})$ conditional on $\mathcal{F}_t$ can be written as

$$\begin{aligned} \psi(r_1, r_2; x_{t+\tau}, v_{t+\tau} \mid x_t, v_t) &= E[\exp\{ir_1 x_{t+\tau} + ir_2 v_{t+\tau}\} \mid x_t, v_t] \\ &= \exp\{C(\tau; r_1, r_2) + D1(\tau; r_1, r_2)' \, x_t + D2(\tau; r_1, r_2)' \, v_t\}, \end{aligned}$$

where $C(\cdot)$, $D1(\cdot)$ and $D2(\cdot)$ are solved from the Ricatti equations as

$$\begin{aligned} C(\tau; r_1, r_2) &= (ir_1\mu + i\beta\alpha r_2)\tau + \frac{\alpha\beta}{\sigma^2} \left[(b - h)\tau - 2 \ln \left(\frac{1 - ge^{-h\tau}}{1 - g} \right) \right] \\ D1(\tau; r_1, r_2) &= ir_1 \\ D2(\tau; r_1, r_2) &= ir_2 + \frac{b - h}{\sigma^2} \frac{1 - e^{-h\tau}}{1 - ge^{-h\tau}}, \end{aligned}$$

with $h(r_1, r_2) = [b^2 + \sigma^2(r_1^2 + 2\rho\sigma r_1 r_2 + \sigma^2 r_2^2 + 2i\beta r_2)]^{1/2}$, $b = \beta - \rho\sigma i r_1 - \sigma^2 r_2 i$, $g(r_1, r_2) = (b - h)/(b + h)$.