ELSEVIER

Journal of Econometrics 102 (2001) 111–141

JOURNAL OF Econometrics

www.elsevier.nl/locate/econbase

Estimation of affine asset pricing models using the empirical characteristic function

Kenneth J. Singleton*

Graduate School of Business, Stanford University, Stanford, CA 94305-5015, USA

Received 6 August 1999; received in revised form 1 August 2000; accepted 2 October 2000

Abstract

The known functional form of the conditional characteristic function (CCF) of discretely sampled observations from an affine diffusion is used to develop computationally tractable and asymptotically efficient estimators of the parameters of affine diffusions, and of asset pricing models in which the state vectors follow affine diffusions. Both 'time-domain' estimators, based on Fourier inversion of the CCF, and 'frequency-domain' estimators, based directly on the CCF, are constructed. A method-of-moments estimator based on the CCF is shown to approximate the efficiency of maximum likelihood for affine diffusion and asset pricing models. © 2001 Elsevier Science S.A. All rights reserved.

JEL classification: C13; C22; G12

Keywords: Affine asset pricing; Efficient estimation; Empirical characteristic function

1. Introduction

Econometric analysis of continuous-time, dynamic asset pricing models is computationally challenging, because the implied conditional density functions of discretely sampled returns are the solutions to partial differential equations (PDEs) and, often, the asset prices themselves must also be computed

* Tel.: + 1-650-723-5753; fax: + 1-650-725-9932.
E-mail address: ken@future.stanford.edu (K.J. Singleton).

PII: S0304-4076(00)00092-0

numerically as nonlinear functions of the underlying state variables. Motivated in part by these considerations, considerable attention has recently been focused on affine asset pricing models – models in which the drift and diffusion coefficients of the state process are affine functions – because they lead to closed- or nearly closed-form expressions for certain asset prices.[1] The tractability of pricing in affine models has expanded substantially the class of asset pricing models that have been studied econometrically. Yet, outside of the special cases of Gaussian and square-root diffusions,[2] where the conditional densities of the discretely sampled returns are also known in closed-form, maximum likelihood methods remain largely unused. This is evidently because of the apparent need to solve PDEs for the conditional density functions.

This paper exploits the conditional characteristic function (CCF) of discretely sampled observations from an affine diffusion to develop computationally tractable and asymptotically efficient estimators of the parameters of affine diffusions, and of asset pricing models in which the state vectors follow affine diffusions. The key observation underlying the proposed estimation strategies is that, if $\{Y_t\}$ is a discretely sampled time series from an N-dimensional affine diffusion, then the CCF of Y_{t+1}, conditioned on Y_t – denoted by $\phi_t(u,\gamma)$ for constant real u and vector of model parameters γ – is known in closed form as an exponential of an affine function of Y_t (see Duffie, Pan, and Singleton, 2000 and Section 2). We use this observation to develop several 'time domain' estimators based on Fourier inversion of $\phi_t(u,\gamma)$, which gives the conditional density function of Y_{t+1} given Y_t. Additionally, method-of-moments estimators are developed directly in the 'frequency domain' by exploiting the fact that $E[e^{iu'Y_{t+1}}|Y_t] = \phi_t(u,\gamma)$, for imaginary i. These 'empirical CCF' estimators avoid the need for Fourier inversion.

We address two related estimation problems using the CCF: (AD) discretely sampled observations on an affine diffusion Y_t are available for estimation of the parameter vector γ_0 governing the conditional distribution of Y_{t+1} given Y_t; and (APAD) the vector of observed prices/yields y_t is described by an affine asset pricing model as $y_t = \mathscr{P}(Y_t, \gamma_0)$, where the unobserved state process Y_t follows an affine diffusion, γ_0 includes the parameters governing the conditional distribution of the state Y_t and those of the pricing model $\mathscr{P}$, and $\mathscr{P}$ is an affine function of Y_t. A third case that is as in case APAD, except that $\mathscr{P}$ is a nonlinear function of Y_t, is briefly discussed in Section 3 and in the concluding remarks.

[1] See Duffie and Kan (1996) and Dai and Singleton (2000) and the references therein for discussions of affine models of term structures of bond prices, Bates (1997) and Bakshi et al. (1997) for models of option prices in which equity returns follow affine, stochastic volatility models, and Backus et al. (1996) and Bates (1996) for discussions of affine models of foreign currency exchange rates.

[2] In the term structure literature, these cases are often referred to as the 'Vasicek' and 'CIR' models.

Estimation problem AD arises in descriptive studies of asset returns. Problem APAD applies, for instance, to affine term structure models where the data is comprised of yields on zero-coupon bond (Duffie and Kan, 1996; Dai and Singleton, 2000). In both of these cases, the functional form of the CCF of the data is known in closed form based on knowledge of the CCF of the state process Y_t, $\phi_t(u, \gamma)$.

When the number of observed prices/yields in y_t is the same as the dimension of the state Y_t, fully efficient ML estimators of γ_0 can be computed for both estimation problems (as well as the case of nonlinear pricing) from knowledge of $\phi_t(u, \gamma)$. In Section 3 we use the CCF *inversion formula* to derive the pricing model-implied conditional log-likelihood function for y_{t+1} given y_t. Maximizing this likelihood function gives the asymptotically efficient 'ML-CCF' estimator of γ_0. We illustrate this estimation strategy for the case of discretely sampled data from a square-root diffusion.

Even though the CCF of Y_t is known essentially in closed form, the computational burdens of ML-CCF estimation can grow rapidly as the dimension of Y increases. This is because, when $N \geqslant 2$, multivariate Fourier inversions must be computed repeatedly and accurately to maximize the likelihood function. Therefore, we proceed to study several computationally simpler, limited-information estimation strategies even though the CCF of the observed prices y_t is known.

Specifically, in Section 4, we propose the limited-information (LML-CCF) estimator based on the conditional density functions $f(y_{j,t+1}|y_t; \gamma)$ of the individual $y_{j,t+1}$ *conditioned on the entire state vector* y_t. The LML-CCF estimator fully exploits the information in the conditional likelihood function of the individual $y_{j,t+1}$, but not the information in the joint conditional distribution of y_{t+1}. The consequent loss in asymptotic efficiency relative to the ML-CCF estimator is traded off against a potentially large reduction in the computational demands of LML-CCF estimation. Moreover, the LML-CCF estimator is typically more efficient than the quasi-ML (QML) estimator for affine diffusions proposed by Fisher and Gilles (1996), among others.

The CCF can be used, as well, to derive closed-form expressions for the conditional moments of y_{t+1}, given y_t, by evaluating the derivatives of the CCF at zero. By definition, the difference between the jth power of the ith element of y_t, y_{it}^j, and its CCF-implied theoretical counterpart will be mean independent of the conditioning variables. GMM estimators of the unknown parameters, based on these conditional moments, are also discussed in Section 4. Liu (1997) develops a complementary estimator based on direct calculation of the conditional moments of affine diffusions and shows that, as the number of moments included is increased to infinity, this GMM estimator attains the efficiency of the ML estimator.

The LML-CCF and GMM estimators of γ_0 necessarily sacrifice some efficiency for computational tractability. We show in Section 5 that there is an alternative, 'frequency domain' estimator that achieves, approximately, the same efficiency as the ML-CCF estimator. This empirical CCF estimator is constructed directly from the CCF and, thereby, avoids the need for Fourier inversion. The exponential function $e^{iu'Y_{t+1}}$ is evaluated at a finite grid of points $u \in \mathbb{R}^N$, and then an optimal method-of-moments estimator is constructed based on the conditional moment restriction $E[e^{iu'Y_{t+1}}|Y_t] - \phi_t(u, \gamma_0) = 0$ at the *true* population parameter vector γ_0. The asymptotic efficiency of this 'GMM-CCF' estimator is shown to approach that of the ML-CCF estimator as the grid of u's becomes increasingly fine in $\mathbb{R}^N$. Moreover, for any fixed, finite grid in $\mathbb{R}^N$ at which the CCF is evaluated, the GMM-CCF estimator is consistent and its asymptotic covariance matrix is easily computed. Heuristically, the GMM-CCF estimator is the solution to an approximation of the first-order conditions to the frequency domain representation of the log-likelihood function, chosen in such a way that consistency is maintained for any degree of accuracy of the approximation.

To gain some insight into the relative efficiencies of the fully and approximately efficient estimators (ML-CCF and GMM-CCF, respectively), we compare the associated asymptotic covariance matrices for an illustrative square-root diffusion model. For this univariate affine model, evaluation of the empirical CCF at only two points gives a GMM-CCF estimator that closely approximates the efficiency of the ML-CCF estimator.

There are several alternative strategies that have recently been developed for the estimation of diffusions using discretely sampled data. One is the simulated method of moments estimator proposed by Gallant and Tauchen (1996) and Gallant and Long (1997).[3] They approximate the likelihood function of the true data generating process by that of a semi-nonparametric auxiliary model, and use the associated scores to construct an SMM estimator. As their approximate density becomes arbitrarily close to the true conditional density, this method-of-moments estimator approaches the efficiency of the ML estimator. Thus, for the class of models examined in this paper, the ML-CCF estimator differs from the Gallant–Long estimator in that it is exact maximum likelihood estimation, while the GMM-CCF and Gallant–Long estimators are both approximately efficient. Alternative, approximately efficient, time-domain estimators are presented in Pedersen (1995), Duffie, Pedersen, and Singleton (2000), Brandt and Santa-Clara (2001), and Ait-Sahalia (1999).

By exploiting the known CCF of affine diffusions, the estimators proposed here may have computational advantages over simulation-based estimators. Furthermore, the criterion function of the GMM-CCF estimator involves a known distance matrix, thereby avoiding the usual two steps of GMM estimation (Hansen, 1982). This is possible, effectively, because the elements of

[3] Among the applications of this approach to affine models that we are aware of are the study by Dai and Singleton (1996) of affine term structure models, and the study by Andersen et al. (1998) of a stochastic volatility model for stock returns.

the optimal distance matrix in the GMM criterion function are known in closed form as functions of the CCF. Based on the existing Monte Carlo evidence for GMM estimators for other asset pricing settings (e.g., Richardson and Smith, 1991), we suspect that the absence of a need to estimate a distance matrix also has advantages in terms of the small sample properties of the optimal GMM-CCF estimator for affine diffusions.

There is also a large literature on estimation and inference for the marginal distributions of random variables using the characteristic function, mostly for i.i.d. environments.[4] Knight and Satchell (1997) and Knight and Yu (1998) propose using the unconditional CF of $(Y_t, \dots, Y_{t-\ell})$ to estimate the parameters of certain time-series models, including a Gaussian ARMA process Y_t. Feuerverger and McDunnough (1981b) and Feuerverger (1990) discuss the estimation of the parameters of the distribution of (y_{t+1}, y_t) using the joint ECF $e^{i(uy_{t+1}+wy_t)}$ for the case of generic stationary Markov time series. Our complementary estimation strategies for affine diffusion and APAD models achieve the asymptotic efficiency of the ML estimator (actually or approximately) by exploiting knowledge of the CCF of the distribution of y_{t+1} given y_t. Finally, Chacko and Viceira (1999b) independently propose, for certain continuous-time models, an inefficient version of our GMM-CCF estimator, and Das (2000) uses the CCF for the special case of Poisson–Gaussian affine diffusions to compute conditional moments of interest rates.

2. Affine diffusions, pricing models, and CFs

This section defines the affine diffusion process and associated affine asset pricing models that will be the focus of this analysis, derives the CCF for an affine diffusion, and outlines the regularity conditions that will be maintained throughout.

2.1. Affine diffusions

For a given complete probability space $(\Omega, \mathscr{F}, P)$ and the augmented filtration $\{\mathscr{F}_t: t \geqslant 0\}$ generated by a standard Brownian motion W in $\mathbb{R}^N$, we suppose there is a Markov process Y taking values in some open subset D of $\mathbb{R}^N$ and satisfying the stochastic differential equation,

$$dY_t = \mu(Y_t, t)\,dt + \sigma(Y_t, t)\,dW_t, \tag{1}$$

[4] See for example Paulson et al. (1975) and Madan and Seneta (1987) for applications to the estimation of the distribution of (presumed i.i.d.) stock return processes, and Heathcote (1972) and Epps et al. (1982) for applications of the ECF and empirical moment generating functions, respectively, to inference.

where $\mu: D \to \mathbb{R}^N$ and $\sigma: D \to \mathbb{R}^{N\times N}$ are regular enough for (1) to have a unique (strong) solution. The Y's may represent, for example, observed asset returns or prices as in descriptive studies, or unobserved state variables in a dynamic pricing model as in affine term structure models.

The diffusion for Y is 'affine' if

$$\mu(y) = \theta + \mathscr{K} y$$

$$\sigma(y)\sigma(y)' = h + \sum_{j=1}^{N} y_j H^{(j)}, \tag{2}$$

where θ is $N \times 1$, $\mathscr{K}$ is $N \times N$, and h and $H^{(j)}$ (for $j = 1, \dots, N$) are all $N \times N$ and symmetric. Duffie and Kan (1996) and Dai and Singleton (2000) discuss conditions on the domain D and the coefficients of μ and $\sigma\sigma'$ under which there is a unique (strong) solution to the SDE (1).

2.2. Affine asset pricing models

Suppose that asset prices are determined by an $N \times 1$ vector of state variables Y_t that follows an affine diffusion. By an affine pricing model we will mean that the instantaneous discount rate r_t at date t is an affine function of the state:

$$r_t = \delta_0 + \delta_y' Y_t, \tag{3}$$

and the payoffs on securities are functions $g(Y_T)$ of the state, so that risk-neutral pricing gives

$$P_t^T = \mathrm{E}_t^q\left[\mathrm{e}^{-\int_t^T r_s \, \mathrm{d}s} g(Y_T) \right], \tag{4}$$

where E^q denote expectation under the risk-neutral measure. The functions g need not be, and generally will not be, affine functions.

The affine term structure models studied, for example, in Duffie and Kan (1996) and Dai and Singleton (2000) are obtained as a special case of (4) by setting $g(Y_T) = 1$, in which case P_t^T is the price of a $(T - t)$-period zero-coupon bond. Similarly, the affine currency pricing models examined in Brandt and Santa-Clara (2001) and Backus et al. (1996) are also special cases for suitably chosen g. Alternatively, suppose that the logarithm of a common stock price is described by $\log S_t = \eta_0 + \eta_y' Y_t$ and $g(Y_T) = \max(S_T - K, 0)$, for given strike price K. Then (4) is an affine option pricing model that includes the models studies by Heston (1993) and the large literature building upon his formulation (see Section 6 for further discussion of this model).

We let γ_0 denote the $Q \times 1$ vector of unknown parameters governing $\mu(y)$, $\sigma(y)$, and the parameters (if any) introduced through an affine pricing model. The latter would include δ_0 and δ_y in (3), as well as the parameters describing the market prices of risk associated with Y. We let Θ denote the admissible parameter space, and assume that it is compact.

2.3. CCFs *of affine diffusions*

The CCF of the Markov process Y_T, conditioned on current and lagged information about Y at date t, is

$$\phi_t(\tau, u) \equiv \mathrm{E}(\mathrm{e}^{\mathrm{i}u'Y_T} | Y_t), \quad u \in \mathbb{R}^N, \tag{5}$$

where $\tau = (T - t)$, $\mathrm{i} = \sqrt{-1}$. Duffie, Pan, and Singleton (2000) prove that the affine structure specified in (2) implies, under technical regularity conditions, that $\phi_t(\tau, u)$ has the exponential-affine form:[5]

$$\phi_t(\tau, u) = \mathrm{e}^{\alpha_t(u) + \beta_t(u)'Y_t}, \tag{6}$$

with α and β satisfying the complex-valued Riccati equations,[6]

$$\dot{\beta}_t = -K'\beta_t - \tfrac{1}{2}\beta_t' H \beta_t, \tag{7}$$

$$\dot{\alpha}_t = -\theta \cdot \beta_t - \tfrac{1}{2}\beta_t' h \beta_t, \tag{8}$$

with boundary conditions $\beta_T(u) = u$ and $\alpha_T(u) = 0$.

The case we will focus on is that of $\tau = 1$, where time is measured in units of the sampling interval of the available data, so that ϕ is the characteristic function of Y_{t+1} conditioned on Y_t. In this case, we suppress the dependence of ϕ on τ and simply write $\phi_t(u)$. Adaptation of the proposed estimators to the case of $\tau > 1$ is immediate. To highlight the dependence of the conditional CF on the unknown parameter vector γ, we will write $\phi_t(u, \gamma)$.

2.4. *Extensions to include jumps*

Though we will focus on affine diffusions, virtually all of the subsequent discussion extends immediately to the case of affine jump diffusions

$$\mathrm{d}Y_t = \mu(Y_t)\mathrm{d}t + \sigma(Y_t)\mathrm{d}W_t + \mathrm{d}Z_t, \tag{9}$$

where Z is a pure jump process with intensity $\{\lambda(Y_t)\colon t \geqslant 0\}$ and jump amplitude distribution ν on $\mathbb{R}^N$. Duffie, Pan, and Singleton (2000) show that the CCF of affine jump diffusions are also known in closed form. Specifically, if the jump intensity is an affine function of Y_t, $\lambda(Y_t) = l_0 + l_y' Y_t$, and the 'jump transform' $\varphi(c) = \int_{\mathbb{R}^N} \exp(c \cdot z) \mathrm{d}\nu(z)$, for $c \in \mathbb{C}^N$, is known in closed form whenever the integral is well defined, then the Riccati equations defining the CCF

[5] Duffie, Pan, and Singleton (2000) prove a more general result that applies to time-dependent coefficients of the diffusion (1).

[6] Here, $c^{\mathrm{T}} H c$ denotes the vector in $\mathbb{C}^n$ with kth element $\sum_{i,j} c_i (H)_{ijk} c_j$.

of Y have

$$\dot{\beta}_t = -K'\beta_t - \frac{1}{2}\beta_t' H\beta_t - l_0(\varphi(\beta_t) - 1), \tag{10}$$

$$\dot{\alpha}_t = -\theta\cdot\beta_t - \frac{1}{2}\beta_t' h\beta_t - l_y(\varphi(\beta_t) - 1). \tag{11}$$

Examples of jump amplitude distributions with known transforms φ are the normal and exponential distributions. The latter has a non-negative range and is therefore useful for modeling jumps in volatility and other variables that are inherently non-negative.

The most widely studied jump-diffusion model for asset prices is the Poisson–Gaussian model in which Y_t follows a Gaussian process with Poisson jumps. In Ball and Torous (1983), Jorion (1988), and Das (2000) for example, the conditional density function of returns was known in closed form so ML estimation proceeded directly. The ML-CCF and GMM-CCF estimators proposed in this paper allow efficient estimation of the entire class of affine jump-diffusion models.

2.5. Regularity conditions

For the estimators discussed in Sections 3–5, we assume that Hansen (1982)'s regularity conditions are satisfied. For the estimation of NPAD models discussed briefly in Section 6 we maintain the regularity conditions for weak consistency and asymptotic normality of GMM estimators adopted by Duffie and Singleton (1993). Though simulation is not used, the proposed estimation strategy uses the 'model-implied' state variables, which are parameter dependent. The regularity conditions and theorems in Duffie and Singleton (1993) cover this situation as a special case.

3. ML-CCF estimators of affine models

A natural way of exploiting the CCF in estimation is to maximize the log-likelihood function obtained by Fourier inversion of the CCF. We will refer to the resulting estimator as the ML-CCF estimator. For the purposes of both highlighting some of the issues that arise in the estimation of affine asset pricing models and motivating subsequent discussions of CCF-based, limited-information estimators, it is instructive to distinguish between three cases: (i) discretely sampled observations $\{Y_t\}$ are observed directly and γ_0 is the parameter vector governing the conditional distribution of Y_{t+1} given Y_t; (ii) the vector of observed prices/yields $y_t = \mathscr{P}(Y_t)$ is described by an affine asset pricing model and $\mathscr{P}$ is an affine function of Y_t; and (iii) $y_t = \mathscr{P}(Y_t)$ is described by an affine asset pricing model and $\mathscr{P}$ is a nonlinear function of Y_t. Throughout this section, we assume that the dimension of Y_t, N, is the same as the dimension of the observed set of asset prices or returns y_t. The ML-CCF estimation strategy is easily extended to accommodate measurement or pricing errors, as is commonly done in the empirical asset pricing literature when there are more security prices than state variables (see Section 6).

3.1. AD *models*: *Discretely sampled* Y_t

Suppose that $\{Y_t\}_{t=1}^{T}$ represents an observed sample from an affine diffusion representation of asset prices or yields. Let $\phi_{Yt}(u, \gamma)$ denote the known CCF of Y_{t+1} given Y_t, and let γ_0 denote the parameter vector of the data-generating process for Y_t. By definition, $\phi_{Yt}(u, \gamma)$ is the Fourier transform of the density function of Y_{t+1} conditioned on Y_t,

$$\phi_{Yt}(u, \gamma) = \int_{\mathbb{R}^N} f_Y(Y_{t+1} | Y_t; \gamma) e^{iu'Y_{t+1}} \, dY_{t+1}. \tag{12}$$

Therefore, the conditional density function of Y_{t+1} is also known, up to an inverse Fourier transform of $\phi_{Yt}(u, \gamma)$:[7]

$$f_Y(Y_{t+1} | Y_t; \gamma) = \frac{1}{\pi^N} \int_{\mathbb{R}^N_+} \mathrm{Re}[e^{-iu'Y_{t+1}} \phi_{Yt}(u, \gamma)] \, du, \tag{13}$$

where Re denote the real part of complex numbers. Given (13), it follows that the conditional log-likelihood function of the sample $\{Y_t\}_{t=1}^{T}$, $\ell_T(\gamma)$, is given by

$$\ell_T(\gamma) = \frac{1}{T} \sum_{t=1}^{T} \log\left\{ \frac{1}{\pi^N} \int_{\mathbb{R}^N_+} \mathrm{Re}[e^{-iu'Y_{t+1}} \phi_{Yt}(u, \gamma)] \, du \right\}. \tag{14}$$

Maximization of (14) can proceed in the usual way, conjecturing a value for γ, computing the associated Fourier inversions, etc.

To illustrate this estimation strategy, and some of the characteristics of a CCF of an affine diffusion, suppose that the instantaneous short-rate r follows a one-factor square-root diffusion process:

$$dr = \kappa(\theta - r)\, dt + \sigma\sqrt{r}\, dB_r. \tag{15}$$

Cox et al. (1985) show that the distribution of $r_{t+\Delta}$ conditioned on r_t is non central $\chi^2[2cr_t, 2q+2, 2\lambda_t]$, where $c = 2\kappa/(\sigma^2(1 - e^{-\kappa\Delta}))$, $\lambda_t = cr_t e^{-\kappa\Delta}$, $q = 2\kappa\theta/\sigma^2 - 1$, and the second and third arguments are the degrees of freedom

[7] R^N_+ is the subspace of R^N with all elements of $u \in R^N$ being non-negative.

and noncentrality parameters, respectively. It follows that the conditional characteristic function for $r_{t+\Delta}$ is

$$\phi_{rt}(u) = (1 - \mathrm{i}u/c)^{-2\kappa\Delta\theta/\sigma^2} \exp\left\{\frac{\mathrm{i}u\mathrm{e}^{-\kappa\Delta}r_t}{(1 - u\mathrm{i}/c)}\right\}. \tag{16}$$

For illustrative purposes, we set the parameter values at $\kappa = 0.4$, $\theta = 6.0$, $\sigma = 0.3$, and $\Delta = 1$ for weekly data. These values are similar to what would be obtained from fitting a square-root diffusion model to weekly data on a short-term interest rate series during a sample period when the average annualized short rate is about 6.0%.[8]

From (14) we see that computation of the likelihood function requires integration only over the real part of $\mathrm{e}^{-\mathrm{i}ur_{t+1}}\phi_{rt}(u)$, which is displayed in Fig. 1 evaluated at the points $(r_{t+1}, r_t) = (6.3, 6.0)$. Being a weighted sum of cosines, this integrand exhibits much less oscillatory behavior than $\phi_{rt}(u)$ itself. Furthermore, the oscillatory behavior in Fig. 1 has largely damped out at about $u = 30$, so truncating the integral in (14) at just over this value would, in this case, give a reasonable approximation to the likelihood function. The degree of oscillation in these functions and their phase relative to each other depends on the distance between r_{t+1} and r_t and whether r_t is above or below its long-run mean (6.0 in this case). Given (r_{t+1}, r_t), the more volatile is r, the more oscillatory is the integrand in the computation of the likelihood function.

To implement the ML-CCF estimator, with $\{r_t\}$ treated as an observed process, we generated one thousand weekly observations by simulation of (15) using an Euler approximation to the diffusion with 50 discrete steps between each weekly observation.[9] The conditional density of r_{t+1} given r_t was computed by Gauss–Legendre quadrature, with various numbers of quadrature points q_p in the approximation to the integral. The ML estimates and their standard errors (in parentheses) are displayed in Table 1. When q_p is at least as large as 20, virtually identical ML-CCF estimators are obtained as q_p is increased. These results are encouraging in that a quite small number of quadrature points recovers the ML estimates.

Even $q_p = 20$ can lead to a computationally demanding estimation problem in multivariate settings, however. Using the basic product rule, the number of points in the grid for approximating the Fourier inversion increases with $(q_p)^N$, where N is the dimension of Y_t. With these potential computational burdens in

[8] Chen and Scott (1993) estimated a one-factor model for U.S. treasury data and obtained comparable values of κ and θ, but a smaller value of σ. Lowering σ, holding κ and θ fixed, tends to slow the rate at which the CCF damps to zero with increasing u and, hence, increases the range over which the CCF must be integrated to obtain the conditional density.

[9] For this example, we could have sampled directly from the conditional distribution (noncentral chi-square) of the discretely sampled data. No precision was lost in this case by using the Euler approximation, as would commonly be done for other diffusion models.

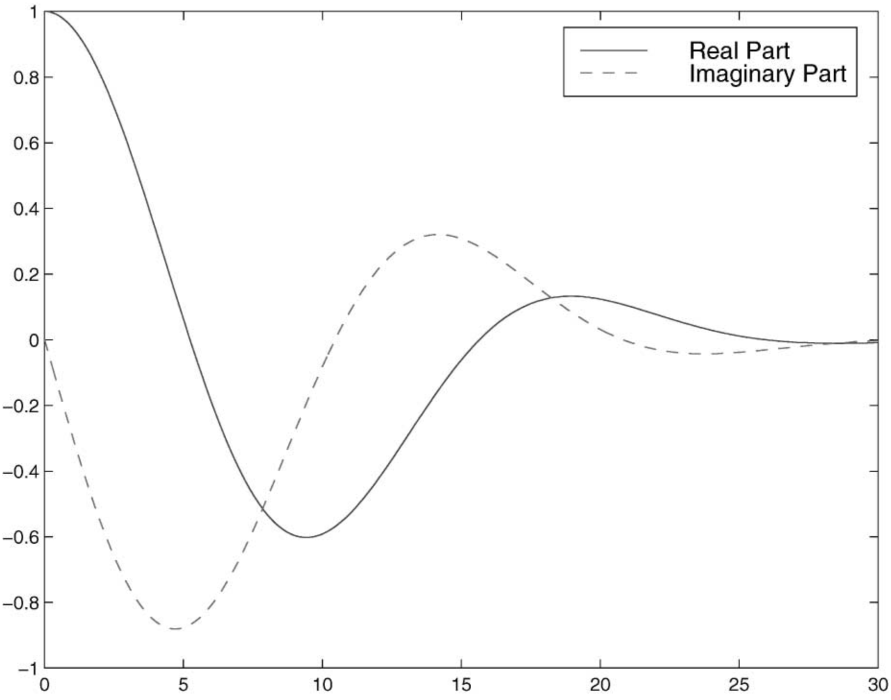

Fig. 1. Plot of real and imaginary parts of integrand for computing conditional density of r evaluated at $(r_{t+1}, r_t) = (6.3,\ 6.0)$.

Table 1
ML-CCF estimates of interest rate model[a]

	κ	θ	σ
Population	0.4	6.0	0.3
$q_p = 10$	0.889	5.930	0.303
	(0.32)	(0.19)	(0.007)
$q_p = 20$	0.377	5.621	0.302
	(0.20)	(0.46)	(0.007)
$q_p = 50$	0.377	5.621	0.302
	(0.20)	(0.46)	(0.007)

[a]Estimated standard errors computed from the Hessian of the likelihood function are given in parentheses.

mind, we explore several less efficient, but computationally less demanding, 'time domain' CCF-based estimators in Section 4, and an approximately efficient empirical CCF estimator in Section 5.

3.2. APAD *models*: *Pricing with affine* $\mathscr{P}(Y_t, \gamma)$

Suppose that the pricing environment is such that $y_t = a(\gamma_0) + B(\gamma_0)Y_t$, with Y_t following an affine diffusion, where the $N \times 1$ vector a and $N \times N$, full-rank matrix B are determined by an affine pricing model. The parameter vector γ_0 includes the parameters governing the affine diffusion Y_t as well as any new parameters introduced by the pricing model. Important special cases of APAD models are affine term structure models in which the short rate r_t is an affine function of an AD Y_t and y_t consists of observations on the yields on zero-coupon bonds (Duffie and Kan, 1996; Dai and Singleton, 2000). In the case of term structure models, γ_0 includes the parameters relating r_t to Y_t in (3) as well as the N market prices of risk associated with Y_t. We will henceforth assume that the parameters of both the affine diffusion and any additional parameters introduced through $\mathscr{P}$ are identified by the moment equations used in estimation. See Dai and Singleton (2000) for a discussion of identification of the parameters in affine term structure models.

Given that y_t is an affine function of Y_t, it follows immediately that[10]

$$\phi_{yt}(u, \gamma) = \mathrm{e}^{\mathrm{i}u'a(\gamma)}\phi_{Yt}(B(\gamma)u), \tag{17}$$

where it is understood that ϕ_Y is evaluated at $Y_t = B(\gamma)^{-1}(y_t - \alpha(\gamma))$. Thus, knowledge of the CCF of Y implies knowledge of the CCF of y and ML-CCF estimation of APAD models can be implemented directly using (17).

In particular, if r_t follows a scalar, square-root diffusion, then the yield on an n-year zero-coupon bond, y_t^n, can be expressed as $y_t^n = a_n(\gamma_0) + b_n(\gamma_0)r_t$, where a_n and b_n are known functions of γ_0. Implicit in the weights a_n and b_n is the dependence of bond prices on the market price of risk λ associated with the state variable r_t (see, e.g., Cox et al., 1985), so $\gamma_0' \equiv (\kappa, \theta, \sigma, \lambda)$. Therefore, using (16), the characteristic function for y_{t+1}^n conditioned on y_t^n is

$$\phi_{y^n t}(u) = \mathrm{e}^{\{-\mathrm{i}ua_n\}}(1 - \mathrm{i}b_n u/c)^{-\kappa\theta 2/\sigma^2}\exp\left\{\frac{\mathrm{i}b_n u\mathrm{e}^{-\kappa}(y_t^n - a_n)/b_n}{(1 - b_n u\mathrm{i}/c)}\right\}. \tag{18}$$

Inversion of this CCF gives the conditional density function for the discretely sampled y_t^n for use in computing ML-CCF estimates of γ_0.

3.3. NPAD *models*: *Pricing with nonlinear* $\mathscr{P}(Y_t, \gamma)$

In the cases of coupon bonds, call options, and other pricing problems, the pricing function $\mathscr{P}(Y_t)$ will be nonlinear and the CCF of the observed prices or yields, y_t, is not known. Nevertheless, the ML-CCF estimator can be

[10] $\phi_{yt}(u, \gamma) = \mathrm{E}[\mathrm{e}^{\mathrm{i}u'y_{t+1}}|y_t] = \mathrm{e}^{\mathrm{i}u'a(\gamma)}\mathrm{E}[\mathrm{e}^{\mathrm{i}u'B(\gamma)Y_t}|y_t] = \mathrm{e}^{\mathrm{i}u'a(\gamma)}\phi_{Yt}(B(\gamma)u)$.

implemented using the standard Jacobian of the transformation $\mathscr{P}$. Assuming that the dimension of y_t is equal to that of Y_t and $\mathscr{P}$ is invertible,

$$f_y(y_{t+1}|y_t;\, \gamma) = f_Y(\mathscr{P}^{-1}(y_{t+1};\, \gamma)|y_t;\, \gamma)\mathrm{abs}\left|\frac{\partial \mathscr{P}^{-1}(y_{t+1};\, \gamma)}{\partial y}\right|. \tag{19}$$

For instance, suppose a researcher has data on the yields on coupon bonds. Letting c_t^n denote the coupon yield on an n-year coupon-paying bond and P_t^n denote the price of an n-year zero-coupon bond, the coupon rate c_t^n for a newly issued n-year bond trading at par is $c_t^n = (100 - P_t^n)/(\sum_{j=1}^{2n} P_t^{0.5j})$, where coupons are assumed to be paid semi-annually. Though each zero price P_t^j is an exponential-affine function of the state, c_t^n is not. However, if $c_t^n = \mathscr{P}(Y_t,\, \gamma_0)$ and $\mathscr{P}$ is invertible so that $Y_t = \mathscr{P}^{-1}(c_t^n;\, \gamma_0)$, then (19) applies. Chen and Scott (1993), Pearson and Sun (1994), and Duffie and Singleton (1997) used the same transformation with the known conditional density of Y to compute ML estimators of multi-factor CIR-style models. Our approach generalizes their method to all affine term structure models using the known CCF of r and, indeed, all affine pricing models.

4. Limited-information estimation

As noted previously, the computational burdens of ML-CCF estimation using the CCF increase with N, so limited-information methods may be attractive when $N > 1$. In this section we outline two limited information estimation methods based on the CCF that are in general less demanding computationally than the ML-CCF estimator. These methods are applicable to any estimation problem where the CCF of the observed prices or yields is known.

4.1. LML-CCF *estimation*

Considerable computational savings are achieved by focusing on the conditional density functions of the individual elements of Y. Let ι_j denote the N-dimensional selection vector with 1 in the jth position and zeros elsewhere. Then the density of $y_{j,t+1} = \iota_j \cdot y_{t+1}$ conditioned on the *entire* y_t is the inverse Fourier transform of $\phi_{yt}(\omega\iota_j,\, \gamma)$ viewed as a function of the scalar ω:

$$f_j(y_{j,t+1}|y_t;\, \gamma) = \frac{1}{(2\pi)}\int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}\omega\iota_j' y_{t+1}} \phi_{yt}(\omega\iota_j,\, \gamma)\, \mathrm{d}\omega. \tag{20}$$

Estimation based on the densities (20) involves at most N one-dimensional integrations, instead of one N-dimensional integration. We will refer to such estimators as *partial*-ML or LML-CCF estimators.

The LML-CCF estimator fully exploits the information in the *marginal* conditional densities of the $y_{j,t+1}$ given y_t. Considerations of efficiency recommend the use of the conditional densities for all $N y_{j,t+1}$ in constructing a LML-CCF estimator, when this is computationally feasible. Importantly, in the context of APAD models with N state variables, the N prices/yields y_t must be computed to implement this estimator even if only one conditional density $f(y_{j,t+1}|y_t;\, \gamma)$ is used in estimation. Thus, the added computational burden of using an additional $f_j(y_{j,t+1}|y_t;\, \gamma)$ in estimation is only the associated, univariate Fourier inversion in (20).

More precisely, fixing j, if the APAD model is correctly specified, then

$$\mathrm{E}\left[\frac{\partial \log f_j}{\partial \gamma}\,(y_{j,t+1}|y_t,\gamma_0)\right] = 0 \tag{21}$$

and hence, under regularity, maximization of the LML-CCF objective function

$$\ell_{jT}(\gamma) = \frac{1}{T}\sum_{t=1}^{T} \log\left\{\frac{1}{(2\pi)}\int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}\omega\iota_j' y_{t+1}}\,\phi_{yt}(\omega\iota_j,\, \gamma)\,\mathrm{d}\omega\right\} \tag{22}$$

gives a consistent estimator of γ_0. One of the regularity conditions is that γ_0 is identified from knowledge of the conditional likelihood function of a single j, $f_j(y_{j,t+1}|y_t;\, \gamma)$. This is the case, for instance, in most multi-factor affine term structure models (Dai and Singleton, 2000). The first-order conditions associated with (22) are

$$\frac{\partial \ell_{jT}}{\partial \gamma}\,(\gamma_T) = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{f_j(y_{j,t+1}|y_t,\, \gamma_T)}\,\frac{1}{(2\pi)}\int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}\omega\iota_j' y_{t+1}}\frac{\partial \phi_{yt}}{\partial \gamma}\,(\omega\iota_j,\, \gamma_T)\,\mathrm{d}\omega = 0. \tag{23}$$

These equations can be interpreted as Q moment conditions in the construction of a GMM estimator γ_T of γ_0. That is, letting $G_T(\gamma) \equiv \partial\ell_{jT}(\gamma)/\partial\gamma$, one can solve these Q equations in Q unknowns to get consistent and asymptotically normal estimates of γ_0.

More efficient estimators will, in general, be obtained by exploiting more than one of the conditional densities (20), say for $j = k_1,\, k_2$. In this case, the first-order conditions for each of the univariate 'log-likelihoods' are stacked to obtain

$$G_T(\gamma) \equiv \begin{pmatrix} \partial\ell_{k_1 T}(\gamma)/\partial\gamma \\ \partial\ell_{k_2 T}(\gamma)/\partial\gamma \end{pmatrix}. \tag{24}$$

Then $G_T(\gamma)' W_T^{-1} G_T(\gamma)$ is minimized over γ, for appropriate choice of distance matrix W_T^{-1} (Hansen, 1982). The moment conditions that underlie G_T are martingale difference sequences, by implication of the model, so the optimal choice W_T is a consistent estimator of $\mathrm{E}[\varepsilon_{t+1}\varepsilon'_{t+1}]$, where

$$\varepsilon_{t+1} \equiv \begin{pmatrix} \partial \log f(y_{k_1,t+1}|y_t;\ \gamma_0)/\partial\gamma \\ \partial \log f(y_{k_2,t+1}|y_t;\ \gamma_0)/\partial\gamma \end{pmatrix}. \tag{25}$$

The most efficient LML-CCF estimator, obtained using the scores of the conditional densities $f(y_{k,t+1}|y_t;\ \gamma)$, for all $k = 1, \dots, N$, is constructed similarly.

Though the LML-CCF estimator does not exploit any information about the conditional joint distribution,[11] information about the conditional covariances can be easily incorporated into the estimation by appending moments to the vector ε_{t+1}. For example, for an affine diffusion, the conditional covariance between $y_{j,t+1}$ and $y_{k,t+1}$ is an affine function of y_t with coefficients that are known functions of γ_0, and the conditional first and second moments of y_{t+1} are easily computed in closed form for an arbitrary AD. Thus, letting

$$\eta_{jk,t+1} \equiv (y_{j,t+1} - \mathrm{E}[y_{j,t+1}|y_t])(y_{k,t+1} - \mathrm{E}[y_{k,t+1}|y_t]), \tag{26}$$

we can add terms of the form $\eta_{jk,t+1}(\gamma)h(y_t)$, where $h: \mathbb{R}^N \to \mathbb{R}$, to ε_{t+1}. Again, by construction, the products $\eta_{jk,t+1}(\gamma_0)h(y_t)$ are martingale difference sequences, so the optimal distance matrix is again computed from a consistent estimator of $\mathrm{E}[\varepsilon_{t+1}\varepsilon'_{t+1}]$. Thus, the LML-CCF estimator potentially embodies a substantial amount of information about the joint distribution of (y_{t+1}, y_t). The costs in terms of asymptotic efficiency loss relative to the ML-CCF estimator may therefore be small relative to the benefits in terms of computational simplicity.

4.2. Conditional moment estimation

The conditional moments of y_{t+1} given y_t can be computed from the derivatives of the CCF evaluated at $u = 0$. Therefore, given a particular conditional moment, say

$$\left.\frac{\partial^{j+k}\phi_{yt}(u,\ \gamma_0)}{\partial u_{s_1}^j\, u_{s_2}^k}\right|_{u=0} = \mathrm{i}^{j+k}\mathrm{E}[y_{s_1,t+1}^j\, y_{s_2,t+1}^k|y_t] \tag{27}$$

for $1 \leqslant s_1,\ s_2 \leqslant N$, orthogonality conditions for GMM estimation can be constructed from the moment restrictions

$$\mathrm{E}\left(y_{s_1,t+1}^j\, y_{s_2,t+1}^k - \left.\frac{\partial^{j+k}\phi_{yt}(u,\ \gamma_0)}{\mathrm{i}^{j+k}\partial u_{s_1}^j\, u_{s_2}^k}\right|_{u=0} \middle| y_t \right) = 0. \tag{28}$$

---

[11] Estimation based on the conditional density functions $f(y_{j,t+1}|y_t)$ does, of course, exploit some information about the correlation among the state variables, since this density is conditional on y_t. In particular, it exploits all of the information about the feedback among the variables through the conditional moments of each $y_{j,t+1}$, $\mathrm{E}[y_{j,t+1}^m|y_t]$.

Similarly, Fisher and Gilles (1996) derived closed-form expressions for the conditional mean $\mathrm{E}[y_{t+1}|y_t]$ and conditional variance $\mathrm{Var}[y_{t+1}|y_t]$, both of which have components that are affine functions of y_t. These moments, which are easily derived from the derivatives of the CCF,

$$\left.\frac{\partial \phi_{yt}(u,\gamma_0)}{\partial u}\right|_{u=0} = \mathrm{iE}[y_{t+1}|y_t]; \quad \left.\frac{\partial^2 \phi_{yt}(u,\gamma_0)}{\partial u \partial u'}\right|_{u=0} = -\,\mathrm{E}[y_{t+1}y'_{t+1}|y_t], \tag{29}$$

can be used to implement a standard QML estimator of γ_0 with the normal likelihood function. This will lead to consistent and asymptotically normal estimators that are generally less efficient than the LML-CCF estimator based on $f(y_{j,t+1}|y_t;\gamma)$, $j=1,\dots,N$.

Outside of the case of Gaussian diffusions, the 'innovations' in affine models are non-normal (e.g., noncentral chi-square in the case of square-root diffusions). The CCF-based estimators exploit information about these nonnormal errors and, thus, in general will be more efficient than the QML estimator. In a different, nonaffine setting, Sandmann and Koopman (1998) found that quasi-ML estimators of stochastic volatility models were relatively inefficient compared to full-information methods, because of the non-normal innovations. One might expect that a similar result would emerge in the case of affine diffusions.

5. ECCF estimation of affine pricing models

All of the estimators discussed in Sections 3 and 4 are 'time-domain' estimators in that they are based directly on conditional densities of y_t. In this section we propose several 'frequency domain' estimators that are based, instead, directly on the CCF. An attractive feature of CCF-based estimators is that they can be constructed to have, approximately, the same asymptotic distribution as the ML-CCF estimator, while being computationally much simpler to implement. In particular, Fourier inversion is not required. We proceed in two steps to derive an asymptotically efficient CCF estimator. First, we derive an asymptotically equivalent, frequency domain representation of the ML-CCF estimator that is the conditional counterpart to a similar characterization of ML estimators for i.i.d environments in Feuerverger and McDunnough (1981b). This estimator turns out to exploit a continuum of conditional moment restrictions involving the CCF. Our proof of its asymptotic efficiency leads directly to the construction in Section 5.2 of an approximately efficient and computationally more attractive estimator that exploits a finite number of these moment restrictions.

5.1. *An efficient* ECCF *estimator*

Using the empirical CCF (ECCF), we begin by constructing an estimator that is asymptotically equivalent to the ML-CCF estimator. Let Z_T^∞ denote the class of 'continuous-grid' ECCF estimators defined as follows. We introduce a set Z_T^∞ of 'instrument' functions with elements $z_t(u):\mathbb{R}^N \to \mathbb{C}^Q$, where $\mathbb{C}$ denotes the complex numbers, with $z_t(u) \in I_t$, $z_t(u) = \bar{z}_t(-u)$, $t = 1,\dots,T$, where I_t is the σ-algebra generated by y_t. Each $z \in Z_T^\infty$ indexes an estimator $\gamma_{\infty T}^z$ of γ_0 satisfying

$$\frac{1}{T}\sum_t \int_{\mathbb{R}^N} z_t(u)[\mathrm{e}^{\mathrm{i}u'y_{t+1}} - \phi_t(u, \gamma_{\infty T}^z)]\,\mathrm{d}u = 0. \tag{30}$$

Under regularity (see Section 2.5), $\gamma_{\infty T}^z$ is consistent, and asymptotically normal with limiting covariance matrix

$$\mathscr{V}_0^\infty(z) = D(z)^{-1}\Sigma^\infty(z)(\bar{D}(z)')^{-1}, \tag{31}$$

where

$$D(z) = \mathrm{E}\left[\int_{\mathbb{R}^N} z_t(u)\,\frac{\partial\phi_t(u)}{\partial\gamma}\,\mathrm{d}u\right], \tag{32}$$

$$\Sigma^\infty(z) = \mathrm{E}\left[\int_{\mathbb{R}^N} z_t(u)[\mathrm{e}^{\mathrm{i}u'y_{t+1}} - \phi_t(u, \gamma_0)]\,\mathrm{d}u \int_{\mathbb{R}^N}[\mathrm{e}^{-\mathrm{i}u'y_{t+1}} - \bar{\phi}_t(u,\gamma_0)]\bar{z}_t(u)'\,\mathrm{d}u\right]. \tag{33}$$

We begin by showing that the optimal index in Z_T^∞, in the sense of giving the smallest asymptotic covariance matrix among continuous-grid ECCF estimators, is

$$z_{\infty t}^*(u) = \frac{1}{(2\pi)^N}\int_{\mathbb{R}^N} \frac{\partial \log f}{\partial\gamma}\,(y_{t+1}|y_t, \gamma_0)'\mathrm{e}^{-\mathrm{i}u'y_{t+1}}\,\mathrm{d}y_{t+1}, \tag{34}$$

and, moreover, the limiting covariance matrix of the GMM estimator $\gamma_{\infty T}^*$ obtained using $z_{\infty t}^*(u)$ is the asymptotic Cramer–Rao lower bound, $I(\gamma_0)^{-1}$. Toward this end, we prove that:[12]

Lemma 5.1. The index $z_{\infty t}^*(u)$ satisfies

$$\int_{\mathbb{R}^N} z_{\infty t}^*(u)\mathrm{e}^{\mathrm{i}u'y_{t+1}}\,\mathrm{d}u = \frac{\partial \log f}{\partial\gamma}\,(y_{t+1}|y_t,\gamma_0)', \tag{35}$$

$$\int_{\mathbb{R}^N} z_{\infty t}^*(u)\phi_t(u,\gamma_0)\,\mathrm{d}u = 0. \tag{36}$$

[12] Proofs are given in Appendix A.

It follows that

$$\int_{\mathbb{R}^N} z^*_{\infty t}(u)[\mathrm{e}^{\mathrm{i}u' y_{t+1}} - \phi_{yt}(u, \gamma_0)]\,\mathrm{d}u = \frac{\partial \log f}{\partial \gamma}(y_{t+1}|y_t, \gamma_0)'. \tag{37}$$

An immediate implication of Lemma 5.1 is that

$$\mathrm{E}\left[\int_{\mathbb{R}^N} z^*_{\infty t}(u)[\mathrm{e}^{\mathrm{i}u' y_{t+1}} - \phi_{yt}(u, \gamma_0)]\,\mathrm{d}u\right] = \mathrm{E}\left[\frac{\partial \log f}{\partial \gamma}(y_{t+1}|y_t, \gamma_0)'\right] = 0, \tag{38}$$

assuming the probability model for y_t is correctly specified. Thus, the sample moments (30) evaluated at $z_t(u) = z^*_{\infty t}(u)$ are asymptotically equivalent to the first-order conditions of the log-likelihood function. It follows that, under regularity, the continuous-grid GMM estimator based on the index $z^*_{\infty t}$ is asymptotically equivalent to the ML estimator based on the true conditional density function of y_t.

5.2. *Approximately efficient* ECCF *estimators*

From a practical perspective, the ECCF estimator $\gamma^*_{\infty T}$ has no computational advantages over the ML-CCF estimator described in Section 3, because the index z^*_∞ cannot be computed without a priori knowledge of the conditional density function. Accordingly, we proceed to develop a computationally tractable estimator that is consistent and 'nearly' as efficient as these ML estimators. For notational simplicity, we set $N = 1$ (y_t is one dimensional) in the remainder of this section.

The basic idea is to approximate the integral

$$\int_{\mathbb{R}} z_t(u)[\mathrm{e}^{\mathrm{i}u y_{t+1}} - \phi_{yt}(u, \gamma)]\,\mathrm{d}u \tag{39}$$

underlying the construction of (30) with the sum over a finite grid in $\mathbb{R}$. For any finite grid, no matter how coarse, this 'GMM-CCF' estimator is shown to be consistent and asymptotically normal with an easily computable asymptotic covariance matrix. Moreover, the asymptotic covariance matrix of the optimal GMM-CCF estimator is shown to converge to $I(\gamma_0)^{-1}$ as the range and fineness of the approximating grid in $\mathbb{R}$ increases.

More precisely, for given $K > 0$ and $\tau > 0$, we fix the interval $[-K\tau, K\tau] \subset \mathbb{R}$ (which is divided into $(2K + 1)$ equally spaced intervals of width τ). Let Z_T^K denote the class of GMM-CCF estimators γ_{KT}^z that solve

$$\frac{1}{T}\sum_t \tau \sum_{k=-K}^{K} z_t(k\tau)[\mathrm{e}^{\mathrm{i}k\tau y_{t+1}} - \phi_{yt}(\tau k, \gamma_{KT}^z)] = 0. \tag{40}$$

This expression, in turn, can be simplified further by letting

$$\begin{aligned}\varepsilon_{K,t+1}(\gamma)' \equiv (&\cos(\tau y_{t+1}) - \mathrm{Re}\,\phi_{yt}(\tau,\gamma), \dots, \cos(K\tau y_{t+1}) - \mathrm{Re}\,\phi_{yt}(K\tau,\gamma),\\ &\sin(\tau y_{t+1}) - \mathrm{Im}\,\phi_{yt}(\tau,\gamma), \dots, \sin(K\tau y_{t+1}) - \mathrm{Im}\,\phi_{yt}(\tau,\gamma)),\end{aligned} \tag{41}$$

and $\tilde{z}_{Kt}$ denote the $Q \times 2K$ real matrix with the $\mathrm{Re}[z_t(k\tau)]$ and $-\mathrm{Im}[z_t(k\tau)]$ being the first K and second K columns, respectively. Then (40) becomes

$$\frac{1}{T}\sum_t \tilde{z}_{Kt}\varepsilon_{K,t+1}(\gamma_{KT}^z) = 0. \tag{42}$$

This estimator is consistent for essentially any $K \geqslant 1$, because each column of $\tilde{z}_{Kt}$ has dimension Q, the number of unknown parameters. The asymptotic distribution of γ_{KT}^z is normal with covariance matrix

$$\mathscr{V}_0^K(\tilde{z}) = \{D^K(\tilde{z})\}^{-1} S_0^K(\tilde{z})\{D^{K\prime}(\tilde{z})\}^{-1}, \tag{43}$$

where

$$S_0^K(\tilde{z}) \equiv \mathrm{E}[\tilde{z}_{Kt}\varepsilon_{K,t+1}(\gamma_0)\varepsilon_{K,t+1}(\gamma_0)' z_{Kt}'], \tag{44}$$

$$D^K(\tilde{z}) \equiv \mathrm{E}\left[\tilde{z}_{Kt}\frac{\partial\varepsilon_{K,t+1}(\gamma_0)'}{\partial\gamma}\right]. \tag{45}$$

Now one of the estimators in Z_T^K has $\tilde{z}_{Kt} = \tilde{z}_{Kt}^{\infty}$, where $\tilde{z}_{Kt}^{\infty}$ is constructed from the real and imaginary parts of $z_{\infty t}^*(k\tau)$. For this choice of instrument function, (40) is a quadrature approximation over the interval $[-K\tau, K\tau]$ to the optimal continuous-grid ECCF estimator presented in the preceding section. Therefore, if τ is chosen as a function of K so that $\tau \to 0$ and $(2K+1)\tau \to \infty$, as $K \to \infty$, then the asymptotic covariance matrix of the fixed-grid estimator $\tilde{z}_{Kt}^{\infty}$ will converge to the asymptotic Cramer–Rao bound: $\lim_{K\to\infty}\mathscr{V}_0^K(z_\infty^*) = I(\gamma_0)^{-1}$.

We exploit this observation to show that there is another, approximately efficient, GMM-CCF estimator that is much more tractable computationally. Applying the results in Hansen (1985), the optimal index $\tilde{z}_{Kt}^* \in Z_T^K$ is[13]

$$\tilde{z}_{Kt}^* = \Phi_t^{K\prime} \times \{\Sigma_t^K\}^{-1}, \tag{46}$$

where the $2K \times Q$ matrix Φ_t^K is $\partial\varepsilon_{K,t+1}(\gamma_0)/\partial\gamma$ and $\Sigma_t^K \equiv \mathrm{E}[\varepsilon_{K,t+1}\varepsilon_{K,t+1}' | y_t]$. Note that the elements of the matrix of derivatives of $\varepsilon_{K,t+1}$ with respect to γ involve only the derivatives of $\mathrm{Re}[\phi_{yt}(k\tau,\gamma)]$ and $\mathrm{Im}[\phi_{yt}(k\tau,\gamma)]$ with respect to γ, and

[13] This estimator will in general not be $\tilde{z}_{Kt}^{\infty}$, because the latter index is only optimal for $K = \infty$, $\tau(K) = 0$.

these terms are in the information set at date t. Moreover, the elements of the matrix Σ_t^K are known in closed form as functions of the real and imaginary parts of $\phi_{yt}(k\tau)$. Thus, $\tilde{z}_{Kt}^*$ is easily computed in practice.

For $\tilde{z}_{Kt} = \tilde{z}_{Kt}^*$, (43) simplifies to

$$\mathscr{V}_0^K(\tilde{z}_K^*) = (\mathrm{E}[\Phi_t^{K\prime}(\Sigma_t^K)^{-1}\Phi_t^K])^{-1}. \tag{47}$$

The optimality property of $\tilde{z}_{Kt}^*$ implies that

$$I(\gamma_0)^{-1} \leqslant \mathscr{V}_0^K(\tilde{z}_K^*) \leqslant \mathscr{V}_0^K(\tilde{z}_K), \tag{48}$$

for any estimator $\tilde{z}_K \in Z_T^K$. With $\tilde{z}_{Kt} = \tilde{z}_{Kt}^\infty$, the right-most term in (48) converges to the left-most term as K approaches ∞ (and τ goes to zero as before). It follows that the optimal GMM-CCF estimator converges to the Cramer–Rao bound as the approximating grid over $u \in \mathbb{R}$ becomes increasingly fine.

5.3. Grid selection in practice

There are several practical considerations that should be kept in mind when selecting a finite grid of u's for the GMM-CCF estimator. One obvious point is that if the conditional distributions of one or more of the y's are symmetric, then the corresponding imaginary parts of the CCF are zero. In this case, the corresponding elements of Φ_t^K and Σ_t^K are omitted.

Identification may also be tenuous in some circumstances, because the CCFs of affine jump diffusions often have periodic or near-period components.[14] Consider, for example, the case of the Poisson distribution that might be present, for example, in jump-diffusion models of asset prices. With $N = 1$, the CF of this distribution is periodic, attaining the value unity at $u_k = 0, \pm 2\pi d, \pm 4\pi d, \ldots$, where d is the distance between lattice points (Lukacs, 1970). Therefore, for any K-vector of u's that are all integral multiples of $2\pi d$, it follows that $|\phi(u_k, \gamma) - \phi(u_k, \hat{\gamma})| = 0$ regardless of the location of γ relative to $\hat{\gamma}$ in the admissible parameter space. While this selection of $\boldsymbol{u}$ leads to an extreme case, whenever any two elements of $\boldsymbol{u}$ differ by an integral multiple of $2\pi d$, the number of columns of $\tilde{z}_{Kt}$ should be reduced by two to avoid a singularity in Σ_t^K.

It is not necessary for the CF to be strictly periodic for this problem to arise. For instance, consider the case of a normal distribution with known variance σ^2 and unknown mean μ_0 and suppose that estimation is based on a single value of u. In this case, the real and imaginary parts of $\phi(u,\gamma)$ are $[\mathrm{e}^{-u^2\sigma^2/2}\cos u\mu_0, \mathrm{e}^{-u^2\sigma^2/2}\sin u\mu_0]$, so there are two orthogonality conditions for use in estimating the single unknown parameter μ_0. However, for $\mu = \mu_0 + 2\pi j/u$ ($j = 1, 2, \ldots$), $\phi(u, \mu) = \phi(u, \mu_0)$ and μ_0 is not identified. Both of these problems are easily

[14] Considerations similar to the following were noted in Epps and Singleton (1982) in their discussion of a goodness-of-fit test for times series based on the ECF.

overcome in practice; e.g., in the second example it is sufficient to use more than one u. However, they highlight the need for some care in selecting $\boldsymbol{u}$ to avoid exact or 'numerical' under-identification of the parameters of an affine model.

Once a set of u's has been selected, numerical problems may arise in computing the GMM-CCF estimator, especially as the grid of u's used to construct $\tilde{z}_{Kt}$ becomes increasingly fine. The matrix Σ_t^K can become ill-conditioned and difficult to invert in actual applications,[15] because some of the $e^{iu_k' y_{t+1}}$, especially for adjacent u's, may be nearly perfectly correlated.

For models with multi-dimensional state vectors ($N > 1$), implementation of the GMM-CCF estimator involves evaluating $\phi_{yt}(u)$ over a grid in a subspace of $\mathbb{R}^N$. As such, the dimension of $\varepsilon_{K,t+1}$, instead of being $2K$, will be $(2K)^N$. Therefore, one may wish to reduce the dimension of $\varepsilon_{K,t+1}$, either for computational reasons or because the conditional covariance matrix of $\varepsilon_{K,t+1}$ becomes nearly singular. This could be accomplished, for example, by constructing a grid of $2K$ points along each axis and then evaluating $\phi_{yt}(u)$ at these points (i.e., for the jth axis, u is a vector of zeros except for the jth entry). The resulting estimator, which is based on a $2KN$-dimensional $\varepsilon_{K,t+1}$, is the frequency domain counterpart of the LML-CCF estimator.

Turning to the efficiency of the GMM-CCF estimator, for fixed K, one can choose $\boldsymbol{u}$ to minimize the asymptotic covariance matrix of γ_{zT}^K. More precisely, after computing a first-stage, consistent estimator of γ_0 using any choice of $\boldsymbol{u}$ that assures consistency, the asymptotic covariance matrix (47) can be minimized as a function of $\boldsymbol{u}$. If one follows the suggestion of Feuerverger and McDunnough (1981a) and selects the elements of $\boldsymbol{u}$ with equal spacing ($u_k = \bar{u} + k\tau$, for fixed τ), then the norm of the covariance matrix can be minimized with respect to the choice of $\bar{u}$ and τ. Alternatively, if equal spacing is not imposed, then one can solve the minimization problem by choice of the entire vector $\boldsymbol{u}$. The latter procedure was suggested by Schmidt (1982) in the context of estimation with a moment generating function.

To explore the relative efficiency of the ECCF estimator proposed here for affine diffusions, we revisit the univariate square-root diffusion model for r in (15) with parameter values given in Table 1. A time series of length 50,000 was simulated from this model using an Euler approximation for the diffusion and then the GMM-CCF estimator was computed for alternative choices of $\boldsymbol{u}$ in the case of $K = 2$. The norms of the asymptotic covariance matrices, trace $[\mathscr{V}_0^2(\tilde{z}_2^*)\mathscr{V}_0^2(\tilde{z}_2^*)']$, for various pairs (u_1, u_2) are displayed in Fig. 2. Interestingly, the norm of $\mathscr{V}_0^2(\tilde{z}_2^*)$ does not vary substantially over a wide range of u's between zero and ten. Having at least one of u_1 or u_2 close to zero does improve estimator efficiency, however.

[15] This problem arose in Madan and Seneta (1987)'s implementation of an empirical CF estimator of i.i.d. stock returns. A similar problem, in a different context, was noted by Carrasco and Florens (1997) in an implementation of a GMM estimator with a continuum of moment conditions.

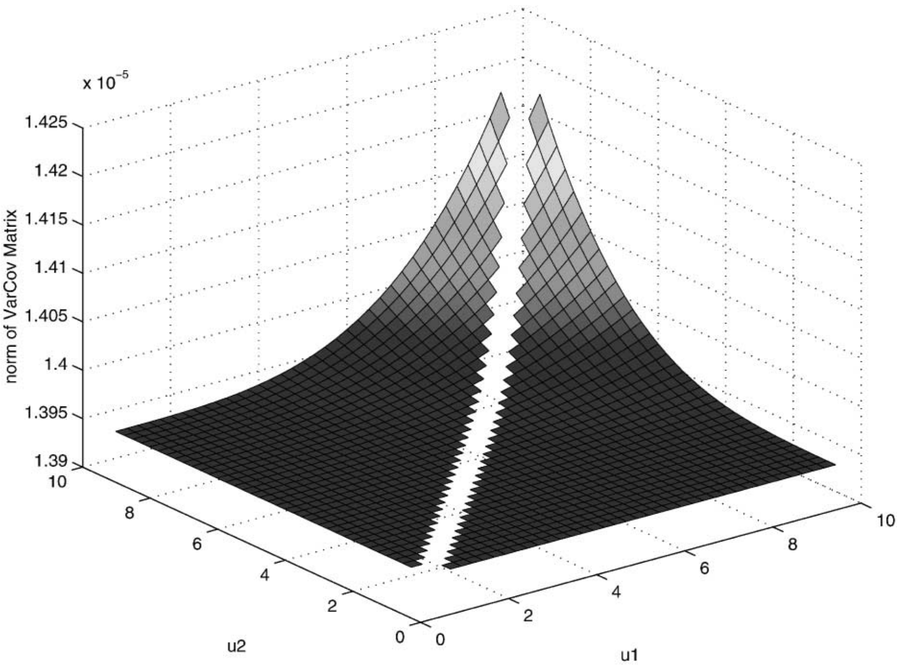

Fig. 2 . Norm of the asymptotic covariance matrices of the ECCF estimator for $K = 2$ and various pairs $(u_1,\ u_2)$.

Examining the individual standard errors at various points on the grid in Fig. 2, we find that they are nearly identical. For instance, setting $u_2 = 0.5$ and $u_1 = 0.75,\ 3.0$, or 10.0, gave virtually identical asymptotic standard errors for all three parameters, and they were identical to those associated with the ML-CCF estimator (i.e., the asymptotic Cramer–Rao bound). These points lie along one of the front axes of Fig. 2 where the norm is smallest. Even at the peak, $(u_1, u_2) = (10.0,\ 10.5)$, the standard errors were the same to three decimal places.[16] Thus, for this model, the asymptotic relative efficiency is high for K set as small as 2 and for a wide range of values of $\boldsymbol{u}$.

More generally, for the study of asset prices, there are a priori reasons for suspecting that the choice of $\boldsymbol{u}$ matters for the large-sample efficiency and small-sample distributions of estimators, and the power of tests. The value of u_k determines the weights given to the moments in the power series expansion of the CCF. Small values of u_k give more weight to the low- than the high-order conditional moments (Lukacs, 1970). We know that many asset returns exhibit conditional skewness and excess kurtosis. Thus, inclusion of large values of

[16] We cannot choose $u_1 = u_2$, because then Σ_t^K would be singular.

u_k may be important for capturing these departures from normality. What is large will depend on the scale of the data, since the sample data enter the ECCF as products with the u_k; i.e., $\cos u_k' y_{t+1}$ and $\sin u_k' y_{t+1}$ for the real and imaginary parts of $e^{iu' y_{t+1}}$. On the other hand, convergence in distribution of γ^z_{KT} may proceed more rapidly for u's concentrated near zero.

6. Concluding remarks

This paper has developed several estimation strategies for affine asset pricing models based on the known functional form of the CCF of affine diffusions. Though our exposition has focused on the diffusion component of the state vector Y_t, as noted in Section 2, all of this discussion extends immediately to a large class of affine jump-diffusion models for Y_t.

A common feature of the affine asset pricing models that have been studied empirically is that the number of available security prices for use in estimation (say M) exceeds, often by a large number, the dimension of Y_t (i.e., $M > N$). One approach to dealing with this difference is to introduce a set of $M - N$ measurement or pricing errors η_{t+1}, and let $y_t = \mathscr{P}(Y_t) + (0', \eta_t')'$, so that the number of sources of uncertainty equals M. This was the approach pursued in the empirical term structure analyses of Chen and Scott (1993) and Dai and Singleton (2000), for example. Assuming that the η process is independent of Y, the CCF of y_{t+1} becomes $\mathrm{E}[e^{iu'\mathscr{P}(Y_{t+1})}|y_t] \times \mathrm{E}[e^{iu'(0', \eta_{t+1}')'}|y_t]$. Given a parametric assumption about the distribution of η_t, this relation can be used to construct ML-CCF estimators of APAD and NPAD models, and GMM-CCF estimators of APAD models.

Throughout this analysis we also presumed that all of the state variables are observed. Letting $y_t' = (y_{1t}', y_{2t}')$, suppose instead that y_{1t} is an N_1 vector of observed variables and y_{2t} is an N_2 ($= N - N_1$) vector of unobserved variables. Partition u conformably as $u' = (u_1', u_2')$. Also, let $\vec{y}_t^{\ell} = \{y_t, y_{t-1}, \ldots, y_{t-\ell}\}$ denote the past ℓ-history of y_t. The CCF of $y_{1,t+1}$, given y_t, is $\phi_{yt}(u_1, 0, \gamma_0)$. In general, even though ϕ_{yt} is evaluated at $u_2 = 0$, the CCF of $y_{1,t+1}$ will depend on the entire vector y_t and, hence, on the unobserved vector y_{2t}. Nevertheless, a CCF-based estimator that uses only the sample of the observed y_{1t} can be constructed. Specifically, consider the CCF of $y_{1,t+1}$ conditioned on $\vec{y}_{1t}^{\ell}$, which can be expressed in terms of $\phi_{yt}(u, \gamma_0)$ as[17]

$$\mathrm{E}[e^{iu_1' y_{1,t+1}}|\vec{y}^{\ell}_{1t}] = \mathrm{E}[\phi_{yt}(u_1, 0, \gamma_0)|\vec{y}^{\ell}_{1t}]. \tag{49}$$

[17] This is an immediate implication of the Markov property of y_t. We have

$$\begin{aligned}\mathrm{E}[e^{iu_1' y_{1,t+1}}|\vec{y}^{\ell}_{1t}] &= \mathrm{E}[\mathrm{E}[e^{iu_1' y_{1,t+1}}|\ \vec{y}^{\ell}_{t}]|\ \vec{y}^{\ell}_{1t}] \\ &= \mathrm{E}[\phi_{yt}(u_1, 0, \gamma_0)|\ \vec{y}^{\ell}_{1t}].\end{aligned}$$

Conditioning is on the history $\vec{y}^{\ell}_{1t}$, instead of y_{1t} alone, because $y_{1,t+1}$ is in general not first-order Markov conditional on its own history.

In rare cases the conditional expectation in (49) will be known in closed form or, if not, one could in principle approximate it using nonparametric methods. As a tractable alternative estimation strategy, we propose to exploit (49) and the law of iterated expectations to construct simulated method-of-moments (SMM-CCF) estimators as follows. Letting $h(\vec{y}^{\ell}_{1t})$ denote any measurable function of $\vec{y}^{\ell}_{1t}$, (49) implies that

$$\mathrm{E}[\mathrm{e}^{\mathrm{i}u_1' y_{1,t+1}} h(\vec{y}^{\ell}_{1t})] = \mathrm{E}[\phi_{yt}(u_1, 0, \gamma_0) h(\vec{y}^{\ell}_{1t})]. \tag{50}$$

Given an 'instrument function' $h(\vec{y}^{\ell}_{1t})$, the left-hand side of (50) is replaced by its sample counterpart, $(1/T)\sum_{t=1}^{T} e^{\mathrm{i}u_1' y_{1,t+1}} h(\vec{y}^{\ell}_{1t})$, which involves only observed variables (y_1's). The right-hand side of (50), on the other hand, is computed by Monte Carlo integration. That is, for a given value of γ, a time series of length $\mathscr{T}$ is simulated from a discretized version of y_t, say $\tilde{y}_t$, and then the population expectation is computed as[18] $(1/\mathscr{T})\sum_{s=1}^{\mathscr{T}} \phi_{\tilde{y}s}(u_1, 0, \gamma) h(\tilde{y}^{\ell}_{1s})$. The differences

$$\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\mathrm{i}u_1' y_{1,t+1}} h(\vec{y}^{\ell}_{1t}) - \frac{1}{\mathscr{T}}\sum_{s=1}^{\mathscr{T}} \phi_{\tilde{y}s}(u_1, 0, \gamma) h(\tilde{y}^{\ell}_{1s}), \tag{51}$$

for various choices of u_1 and instrument functions h, can be used to construct a SMM-CCF estimator of γ_0 by minimizing the SMM criterion function discussed in Duffie and Singleton (1993).[19]

There is an important difference between the information exploited in constructing these SMM estimators and the GMM-CCF estimator. In the latter case, for a given u_1, one constructs *unconditional* moment conditions from the *conditional* moment restriction $\mathrm{E}[\mathrm{e}^{\mathrm{i}u_1' y_{1,t+1}} - \phi_{yt}(u_1, 0, \gamma_0) | \vec{y}_{1t}] = 0$. In contrast, in SMM estimation, one is exploiting knowledge of the *unconditional* moment restrictions (50). There is no associated conditional moment restriction, since the underlying unconditional moments (e.g., the right-hand side of (50)) are computed by Monte Carlo integration. At a practical level, it follows that the 'errors'

---

[18] See Gallant and Long (1997), for example, for a discussion of discretization schemes for use in Monte Carlo simulation of diffusions.

[19] Alternatively, we can use the known functional forms of the conditional moments of $y_{1,t+1}$ implied by the CCF to develop a SMM estimator in the 'time domain'. For instance, an implication of (27), with s_1 and s_2 indexing elements of $y_{1,t+1}$, is that an SMM estimator of γ_0 can be constructed using differences of the form

$$\frac{1}{T}\sum_{t=1}^{T} y^{j}_{s_1,t+1} y^{k}_{s_2,t+1} h(\vec{y}^{\ell}_{1t}) - \frac{1}{\mathscr{T}}\sum_{\tau=1}^{\mathscr{T}} \left.\frac{\partial^{j+k}\phi_{\tilde{y}\tau}(u, \gamma)}{i^{j+k}\partial u^{j}_{s_1} u^{k}_{s_2}}\right|_{u=0} h(\tilde{y}^{\ell}_{1\tau}).$$

(e.g., $[\mathrm{e}^{\mathrm{i}u_1' y_{1,t+1}} h(\vec{y}_{1t}^{\ell}) - \mathrm{E}[\phi_{Yt}(u_1, 0, \gamma_0) h(\vec{y}_{1t}^{\ell})]]$) used to construct the SMM estimators are not martingale difference sequences and the optimal distance matrix will be the spectral density matrix of these errors at the zero frequency (Hansen, 1982). Of course, the reason that the optimal GMM-CCF estimator, for a given grid of u_1's (or its time-domain counterpart) cannot be implemented directly is that the optimal moment conditions involve functions of y_{2t}, and y_{2t} is not observed in this case.

Within the family of affine asset pricing models, the problem of unobserved state variables typically arises in cases where the dimension of the state vector N exceeds the dimension of the vector of observed prices or yields. In the context of affine term structure models, if r_t is an affine function of N state variables and the model is to be estimated with only M $(< N)$ bond yields y_t, then effectively $N - M$ of the state variables will be unobserved. Andersen and Lund (1996) estimate a three-factor model ($N = 3$) of a single short-term interest rate ($M = 1$) using the Gallant–Tauchen SMM approach, for example. The SMM estimators proposed here are alternatives that exploit the special structure of affine term structure models.

Another widely studied example is the class of affine stochastic volatility models for equity returns studied by Heston (1993), Bates (1996, 1997), and Bakshi et al. (1997), among others. A basic version of these models has $x_t \equiv \ln(S_t/S_0)$, where S_t is an equity or currency price, following the process

$$\mathrm{d}x = \mu\,\mathrm{d}t + \sqrt{v}\,\mathrm{d}B_r,$$

$$\mathrm{d}v = \kappa(\theta - v)\,\mathrm{d}t + \sigma\sqrt{v}\,\mathrm{d}B_v, \tag{52}$$

where $\mathrm{d}B_r$ and $\mathrm{d}B_v$ may have nonzero correlation ρ. Heston (1993) and Das and Sundaram (1999) show that the characteristic function for x_{t+1} conditioned on (x_t, v_t), is

$$\phi_{xt}(u) = C(u)\mathrm{e}^{\{\mathrm{i}ux_t + A(u) + B(u)v_t\}}, \tag{53}$$

where $A(u) = \mathrm{i}\mu u$,

$$B(h) = \frac{-u^2[\mathrm{e}^{\psi} - 1]}{(\psi + \gamma)[\mathrm{e}^{\psi} - 1] + 2\psi}, \quad C(u) = \left[\frac{2\psi(\mathrm{e}^{(\psi+\gamma)/2})}{(\psi + \gamma)[\mathrm{e}^{\psi} - 1] + 2\psi}\right]^{2\kappa\theta/\sigma^2}, \tag{54}$$

$\psi = \sqrt{\gamma^2 + \sigma^2 u^2}$ and $\gamma = \kappa - \rho\sigma\mathrm{i}u$. It follows immediately that the CCF of the continuously compounded holding period return $r_{t+1} \equiv x_{t+1} - x_t$, conditioned on (x_t, v_t), is $\phi_{rt}(u) = C(u)\mathrm{e}^{\{A(u) + B(u)v_t\}}$, which depends only on the volatility shock v_t.[20] Thus, the CCF of r_{t+1} conditioned on $\vec{r}_t^{\ell}$ is $\mathrm{E}[\mathrm{e}^{\mathrm{i}u_1 r_{t+1}}|\vec{x}_t^{\ell}] = \mathrm{E}[\phi_{rt}(u_2, 0, \gamma_0)|\vec{x}_t^{\ell}]$.

Turning to the case where the N-vector of observed prices (or yields) are nonlinear functions of an unobserved, N-dimensional state vector Y_t, as in many affine bond and option pricing models, ML-CCF estimation remains feasible by standard change-of-variable arguments. However, the various limited-information estimators for APAD models are not applicable to these nonlinear models, because the CCFs of y_t in the latter models are not known. Yet these CCF-based estimation strategies can be modified to obtain consistent, though relatively inefficient, estimators of nonlinear 'NPAD' models. One strategy is to use the moment equations associated with the first-order conditions of the CCF-based estimators for affine diffusion processes, but with the model-implied state variables $\hat{Y}_t \equiv \mathscr{P}^{-1}(y_t)$ substituted for Y_t. That is, we start with a vector function g, derived from the CCF of an affine diffusion, with the property that $\mathrm{E}[g(Y_{t+1}, Y_t; \gamma)] = 0$ at $\gamma = \gamma_0$, and then base estimation on the sample moments

$$G_T(\gamma) \equiv \frac{1}{T}\sum_{t=1}^{T} g(\hat{Y}_{t+1}^{\gamma}, \hat{Y}_t^{\gamma}; \gamma), \tag{55}$$

where $\hat{Y}_t^{\gamma} = \mathscr{P}^{-1}(y_t; \gamma)$ comes from 'inverting' the pricing model for Y_t as a function of y_t.

When proceeding in this way, care must be taken to preserve legitimate moment equations in the presence of the parameter-dependent $\hat{Y}_t^{\gamma}$. This often requires computation of the first-order conditions of the CCF-based estimators treating Y as known and then replacing Y by $\hat{Y}$ in the resulting first-order conditions.[21] At the same time, when computing the derivative of (55) with

[20] Two recent papers propose related CCF-based estimators of the stochastic volatility model (52). Jiang and Knight (1999) exploit the special structure of this stochastic volatility model to derive the unconditional characteristic function of the vector $x_t \equiv (r_t, r_{t-1}, \ldots, r_{t-\ell})$, for fixed $\ell > 0$, and then minimize an integral over u of a weighted difference between the empirical CF and the theoretical joint (unconditional) CF of x_t. Depending on the choice of weighting function used, their estimator may be more or less efficient than our proposed SMM estimators of model (52). It appears that their estimation strategy is not easily adapted to the entire class of affine models with unobserved states. Chacko and Viceira (1999a) construct a GMM estimator based on the unconditional means of the differences $\mathrm{e}^{\mathrm{i}ur_{t+1}} - \mathrm{E}[\phi_{rt}(u, \gamma_0)|\log S_t]$, for various integer values of u, where the conditional mean $\mathrm{E}[\phi_{rt}(u, \gamma_0)|\log S_t]$ is derived analytically by integrating out the dependence of $\phi_{rt}(u, \gamma_0)$ on v_t. This estimator does not exploit the fact that the preceding difference is orthogonal to all functions of the current and past history of r_t, but it is computationally more tractable than the SMM estimators outlined here.

[21] In particular, the first-order conditions to LML-CCF and QML estimators, obtained after first substituting $\hat{Y}$ for Y into the objective function, will typically not give consistent estimators, because of the parameter dependency of $\hat{Y}$.

respect to γ, the dependence of $\hat{Y}^{\gamma}$ on γ must be taken into account. To see why let $\gamma_0' = (\gamma_{10}', \gamma_{20}')'$, where γ_{20} denotes the parameters governing the affine diffusion representation of Y_t and γ_{10} is the vector of parameters introduced by the NPAD model. Though the conditional density functions of the $Y_{j,t+1}$ do not depend directly on γ_{10} (and, hence, neither do ϕ_{Yt} or $\mu(Y_t;\gamma)$ and $\Sigma(Y_t,\gamma)$), $\hat{Y}^{\gamma}$ does depend on γ_{10}. Hence, so do the moment conditions (55). It is through this indirect dependence that identification of the pricing parameters γ_{10} is achieved in these modified estimation strategies.[22]

In the option pricing literature (e.g., Bakshi et al., 1997), as well as in the financial industry, researchers have often employed a measure of distance between observed and model-implied prices to estimate not only $\hat{Y}$, but also the parameters γ_0 of the model. This approach to estimation, while computationally simple, ignores a substantial amount of information about the structure of affine models that could be used in estimation. The preceding estimation strategy represents one approach to exploiting this information in such a way that we formally obtain consistent estimators with known asymptotic covariance matrices.

Acknowledgements

I would like to thank Qiang Dai, Darrell Duffie, Jun Liu, and Jun Pan for extensive discussions; Andrew Ang, Mark Ferguson, and Yael Hochberg for their thoughtful and careful research assistance, three referees for their constructive comments, and the Financial Research Initiative and Gifford Fong Associates Fund of the Graduate School of Business at Stanford University for financial support.

Appendix A. Efficiency of continuous-grid ECF estimators

This appendix proves that the index

$$\omega_{\infty t}^{*}(u;\gamma_0) = \frac{1}{(2\pi)^N}\int_{-\infty}^{\infty} \frac{\partial \log f}{\partial \gamma}(y|Y_t,\gamma_0)' \mathrm{e}^{-\mathrm{i}u\cdot y}\,\mathrm{d}y \tag{A.1}$$

achieves the asymptotic Cramer–Rao bound.

For any $\gamma \in \Theta$,

$$\int_{\mathbb{R}^N} \omega_{\infty t}^{*}(u,\gamma)\mathrm{e}^{\mathrm{i}u'Y_{t+1}}\,\mathrm{d}u$$

[22] Analogous estimators of option pricing models based on implied volatilities have been implemented by Renault and Touzi (1996) and Pan (2000).

$$= \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} \frac{\partial \log f}{\partial \gamma} (\tilde{Y}_{t+1} | Y_t; \gamma) \, \mathrm{d}\tilde{Y}_{t+1} \times \int_{\mathbb{R}^N} \mathrm{e}^{\mathrm{i}u'(Y_{t+1} - \tilde{Y}_{t+1})} \, \mathrm{d}u$$

$$= \frac{\partial \log f}{\partial \gamma} (Y_{t+1} | Y_t; \gamma). \tag{A.2}$$

Thus, using $\omega^*_{\infty t}(u; \gamma_0)$, we obtain the score of the log-likelihood function evaluated at the true population parameter vector γ_0. Furthermore,

$$\int_{\mathbb{R}^N} \omega^*_{\infty t}(u, \gamma) \phi_t(u) \, \mathrm{d}u$$

$$= \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} \frac{\partial \log f}{\partial \gamma} (\tilde{Y}_{t+1} | Y_t; \gamma) f(\tilde{Y}_{t+1}; \gamma) \int_{\mathbb{R}^N} \mathrm{e}^{\mathrm{i}u'(Y_{t+1} - \tilde{Y}_{t+1})} \, \mathrm{d}u \, \mathrm{d}\tilde{Y}_{t+1} \, \mathrm{d}Y_{t+1}$$

$$= \int_{\mathbb{R}^N} \frac{\partial \log f}{\partial \gamma} (Y_{t+1} | Y_t; \gamma) f(Y_{t+1} | Y_t; \gamma) \, \mathrm{d}Y_{t+1}. \tag{A.3}$$

Evaluating the latter expression at γ_0 gives zero and the conclusion of the lemma follows. □

Using these results, we can prove the asymptotic efficiency of the estimator

$$\frac{1}{T} \sum_t \int_{\mathbb{R}^N} \omega^*_{\infty t}(u; \gamma^\infty_T) [\mathrm{e}^{\mathrm{i}u'Y_{t+1}} - \phi_{Y_t}(u, \gamma^\infty_T)] \, \mathrm{d}u = 0 \tag{A.4}$$

using a standard mean-value expansion. Let

$$h_{t+1}(\gamma^*_{\infty T}) \equiv \int \omega^*_{\infty t}(u; \gamma^*_{\infty T}) [\mathrm{e}^{\mathrm{i}u'Y_{t+1}} - \phi_t(u, \gamma^*_{\infty T})] \, \mathrm{d}u, \tag{A.5}$$

$$H_T(\gamma^*_{\infty T}) \equiv \frac{1}{T} \sum_t h_{t+1}(\gamma^*_{\infty T}). \tag{A.6}$$

A standard mean-value expansion of H_T around γ_0 gives

$$0 = \sqrt{T} H_T(\gamma^*_{\infty T}) = \sqrt{T} H_T(\gamma_0) + \frac{\partial H_T(\gamma^\#_T)}{\partial \gamma} \sqrt{T}(\gamma^*_{\infty T} - \gamma_0), \tag{A.7}$$

where $\gamma^\#_T$ a matrix with columns that satisfy $|\gamma_0 - \gamma^\#_{iT}| \leqslant |\gamma_0 - \gamma^*_{\infty T}|$ and, hence, each column is a consistent estimator of γ_0. By Lemma 5.1,

$$\sqrt{T} H_T(\gamma_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \log f}{\partial \gamma} (y | Y_t, \gamma_0)', \tag{A.8}$$

which is asymptotically normal with covariance matrix $I^{-1}(\gamma_0)$.

Furthermore, from the proof of Lemma 5.1 it follows that

$$H_T(\gamma_T^{\#}) = \frac{1}{T}\sum_t \left[\frac{\partial \log f}{\partial \gamma}(Y_{t+1}|Y_t, \gamma_T^{\#})' + \int_{\mathbb{R}^N} \frac{\partial \log f}{\partial \gamma}(Y_{t+1}|Y_t, \gamma_T^{\#})' f(Y_{t+1}|Y_t; \gamma_T^{\#})\, \mathrm{d}Y_{t+1}\right]. \tag{A.9}$$

Since each column of $\gamma_T^{\#}$ is a consistent estimator of γ_0, the last term in (A.9) converges almost surely to zero as $T \to \infty$. Therefore,

$$\lim_{T\to\infty} \frac{\partial H_T(\gamma_T^{\#})}{\partial \gamma} = \mathrm{E}\left[\frac{\partial \log f^2}{\partial \gamma \partial \gamma'}(Y_{t+1}|Y_t, \gamma_0)\right] = I(\gamma_0), \text{ almost surely.} \tag{A.10}$$

Combining these observations, we have

$$\sqrt{T}(\gamma_{\infty T}^{*} - \gamma_0) \approx I^{-1}(\gamma_0)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \log f}{\partial \gamma}(Y_{t+1}|Y_t, \gamma_0)', \tag{A.11}$$

which converges in distribution to a $\mathrm{N}(0, I^{-1}(\gamma_0))$ random vector.

References

Ait-Sahalia, Y., 1999. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approach. Working Paper, Princeton University.

Andersen, T., Benzoni, L., Lund, J., 1998. Estimating jump-diffusions for equity returns. Working paper, November.

Andersen, T., Lund, J., 1996. Stochastic volatility and mean drift in the short term interest rate diffusion: sources of steepness, level and curvature in the yield curve. Working paper, February.

Backus, D., Foresi, S., Telmer, C., 1996. Affine models of currency prices. Working paper, New York University.

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

Ball, C., Torous, W., 1983. A simplified jump process for common stock returns. Journal of Financial and Quantitative Analysis 18, 53–65.

Bates, D., 1996. Jumps and stochastic volatility: exchange rate processes implicit in PHLX deutschemake options. Review of Financial Studies 9, 69–107.

Bates, D., 1997. Post-'87 Crash Fears in S&P 500 Futures Options. NBER Working paper.

Brandt, M., Santa-Clara, P., 2001. Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. Working paper, Wharton School.

Carrasco, M., Florens, J., 1997. Generalization of GMM to a continuum of moment conditions. Working paper, Ohio State University.

Chacko, G., Viceira, L., 1999a. Dynamic consumption and portfolio choice with stochastic voltality. Working paper, Harvard University.

Chacko, G., Viceira, L., 1999b. Spectral GMM estimation of continuous-time models. Working paper, Harvard University.

Chen, R., Scott, L., 1993. Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. Journal of Fixed Income 3, 14–31.

Cox, J., Ingersoll, J., Ross, S., 1985. A theory of the term structure of interest rates. Econometrica 53 (2), 385–407.

Dai, Q., Singleton, K., 2000. Specification analysis of affine term structure models. Journal of Finance LV, 1943–1978.

Das, S., 2000. The surprise element: jumps in interest rate diffusions. Working paper, Harvard Business School.

Das, S., Sundaram, R., 1999. Of smiles and smirks: a term structure perspective. Journal of Financial and Quantitative Analysis 34, 211–239.

Duffie, D., Kan, R., 1996. A yield-factor model of interest rates. Mathematical Finance 6, 379–406.

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

Duffie, D., Pedersen, L., Singleton, K., 2000. Modeling credit spreads on sovereign debt. Working paper, Stanford University.

Duffie, D., Singleton, K., 1993. Simulated moments estimation of Markov models of asset prices. Econometrica 61, 929–952.

Duffie, D., Singleton, K., 1997. An econometric model of the term structure of interest rate swap yields. Journal of Finance 52, 1287–1321.

Epps, T., Singleton, K., 1982. A goodness of fit test for time series based on the empirical characteristic function. Working paper.

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

Feuerverger, A., McDunnough, P., 1981a. On some Fourier methods for inference. Journal of the American Statistical Association 76, 379–387.

Feuerverger, A., McDunnough, P., 1981b. On the efficiency of empirical characteristic function procedures. Journal of the Royal Statistical Society, Series B 43, 20–27.

Fisher, M., Gilles, C., 1996. Estimating exponential affine models of the term structure. Working paper.

Gallant, A.R., Long, J.R., 1997. Estimation stochastic differential equation efficiently by minimum Chi-square. Biometrika 84, 125–141.

Gallant, A.R., Tauchen, G., 1996. Which moments to match? Econometric Theory 12, 657–681.

Hansen, L., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hansen, L., 1985. A method of calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. Journal of Econometrics 30, 203–238.

Heathcote, C., 1972. A test of goodness of fit for symmetric random variables. Australian Journal of Statistics 14, 172–181.

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

Jiang, G., Knight, J., 1999. Efficient estimation of the continuous time stochastic volatility model via the empirical characteristic function. Working paper, University of Western Ontario.

Jorion, P., 1988. On jump processes in the foreign exchange and stock markets. Review of Financial Studies 1, 427–445.

Knight, J., Satchell, S., 1997. The cummulant generating function estimation method. Econometric Theory 170–184.

Knight, J., Yu, J., 1998. Empirical characteristic function in time series estimation. Working paper, University of Western Ontario.

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

Lukacs, E., 1970. Characteristic Functions. Griffin, London.

Madan, D., Seneta, E., 1987. Simulation of estimates using the empirical characteristic function. International Statistical Review 55, 153–161.

Pan, J., 2000. "Integrated" time series analysis of spot and options prices. Working paper, MIT.

Paulson, A., Halcomb, E., Leitch, R., 1975. The estimation of the parameters of stable laws. Biometrica 62, 163–170.

Pearson, N.D., Sun, T., 1994. Exploiting the conditional density in estimating the term structure: an application to the cox, ingersoll, and ross model. Journal of Finance XLIX (4), 1279–1304.

Pedersen, A., 1995. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. Scandinarian Journal of Statistics 22, 55–71.

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

Richardson, M., Smith, T., 1991. Tests of financial models in the presence of overlapping observations. Review of Financial Studies 4, 227–254.

Sandmann, G., Koopman, S., 1998. Estimation of stochastic volatility models via Monte Carlo maximum likelihood. Journal of Econometrics 87 (2), 271–302.

Schmidt, P., 1982. An improved version of the Quandt-Ramsey MGF estimator for mixtures of normal distributions and switching regressions. Econometrica 50, 501–516.