



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Econometrics 116 (2003) 259–292

JOURNAL OF
Econometrics

www.elsevier.com/locate/econbase

Spectral GMM estimation of continuous-time processes

George Chacko*, Luis M. Viceira

Graduate School of Business Administration, Harvard University, Boston, MA 02163, USA

Abstract

This paper derives a methodology for the estimation of continuous-time stochastic models based on the characteristic function. The estimation method does not require discretization of the stochastic process, and it is simple to apply in practice. The method is essentially generalized method of moments on the complex plane. Hence it shares the efficiency and distribution properties of GMM estimators. We illustrate the method with some applications to relevant estimation problems in continuous-time Finance. We estimate a model of stochastic volatility, a jump–diffusion model with constant volatility and a model that nests both the stochastic volatility model and the jump–diffusion model. We find that negative jumps are important to explain skewness and asymmetry in excess kurtosis of the stock return distribution, while stochastic volatility is important to capture the overall level of this kurtosis. Positive jumps are not statistically significant once we allow for stochastic volatility in the model. We also estimate a non-affine model of stochastic volatility, and find that the power of the diffusion coefficient appears to be between one and two, rather than the value of one-half that leads to the standard affine stochastic volatility model. However, we find that including jumps into this non-affine, stochastic volatility model reduces the power of the diffusion coefficient to one-half. Finally, we offer an explanation for the observation that the estimate of persistence in stochastic volatility increases dramatically as the frequency of the observed data falls based on a multiple factor stochastic volatility model.

© 2003 Published by Elsevier B.V.

JEL classification: G12

Keywords: Continuous-time estimation; Characteristic function GMM; Spectral GMM; Stochastic volatility; Jump–diffusion process; Affine models

* Corresponding author. Tel.: +1-617-495-6884; fax: +1-617-495-6592.

E-mail addresses: gchacko@hbs.edu (G. Chacko), lviceira@hbs.edu (L.M. Viceira).

1. Introduction

Continuous-time mathematics has become one of the essential tools of modern finance. The elegant mathematics of stochastic calculus simplifies the solution of a wide range of important problems in finance. However, while continuous-time models are generally easier to solve than discrete-time models, they are also more difficult to estimate than discrete-time models due to the discrete nature of observable data. Therefore, this paper derives an estimation methodology that expands the set of continuous-time stochastic processes for which estimation without Euler discretization is feasible. We call this estimation technique Spectral GMM. This name emphasizes the fact that this technique is essentially Generalized Method of Moments constructed in a complex (imaginary) setting, as the basis for estimation is the characteristic function of the process.

The use of the characteristic function for parameter estimation has an important precedent in the work of [Feuerverger and McDunnough \(1981a\)](#) and [Feuerverger \(1990\)](#), who have developed characteristic-function based estimation techniques for discrete-time i.i.d. and ARMA processes.¹ Working in continuous-time, [Das \(1996\)](#) and [Bates \(1996\)](#) have also used the characteristic function in estimation problems, but they use it to recover the density function via inversion before estimation. This paper, as well as independent work by [Singleton \(2001\)](#) and [Jiang and Knight \(2002\)](#), shows that this inversion is not necessary, and we can use the characteristic function directly for estimation.

The approach in this paper, as well as in [Singleton \(2001\)](#) and [Jiang and Knight \(2002\)](#), traces its roots to [Lo's \(1988\)](#) method for estimating continuous-time models by maximum likelihood. The key insight in this paper is to perform estimation using the conditional characteristic function of the continuous-time process rather than its conditional density function. The characteristic function solves the same Kolmogorov forward and backward equations as the conditional density. However, the boundary conditions are different for the characteristic function, rendering the solution to the characteristic function a more tractable problem. We show that estimation can be accomplished directly off the characteristic function using spectral moments, via [Hansen's \(1982\)](#) GMM, rather than inverting the characteristic function to recover the density function. Accordingly we refer to this estimation technique as simply Spectral GMM.

The Spectral GMM estimation procedure has several important advantages. First, no discretization of the continuous-time process is necessary for a wide class of relevant univariate and multivariate continuous-time processes; second, this procedure can easily handle certain continuous-time latent variable models, such as the affine stochastic volatility model, because the latent variable can be integrated out of the characteristic function trivially; third, jump processes are no more difficult to estimate than pure diffusion processes using this approach. We demonstrate the versatility of our approach by estimating two stochastic volatility models, a jump–diffusion model and a mixed

¹ We thank Ken Singleton for pointing out to us the statistical literature on the empirical characteristic function.

stochastic-volatility, jump–diffusion model.² These are all models for which the conditional density functions are unknown and are difficult to handle by other methods.

We have already mentioned that Singleton (2001) and Jiang and Knight (2002) have also suggested in independent work the use of characteristic functions for the direct estimation of continuous-time processes. The main difference between the procedure we suggest in this paper and those used in Singleton (2001) and Jiang and Knight (2002) occurs when estimating models with latent variables—such as models with stochastic volatility, or models with time-varying expected returns. In order to take full advantage of conditioning information and achieve asymptotically efficient estimates, Singleton (2001) integrates out the latent variable in the conditional characteristic function. However, because the marginal stock return process in these latent variable models are non-Markov, integration requires using a simulation procedure, which almost always involves discretizing the model. This discretization induces an estimation bias, though it is likely that this bias is much smaller than the bias introduced by simply discretizing the continuous-time model for the entire estimation procedure. The procedure suggested in this paper utilizes the unconditional characteristic function for estimation. This has the advantage of not requiring discretization for path simulation, and of being computationally far less demanding. However, this comes at the cost of efficiency in estimation. When the conditional characteristic function is known in closed form, as it is with the general class of affine processes used commonly in finance, the implementation of our technique is particularly simple.

Jiang and Knight (2002) suggest an estimation technique that lies in between our technique and that of Singleton (2001). They condition on a small part of the data rather than the entire sample path as in Singleton (2001). Then they integrate out the joint process of the latent variable over a time block from the characteristic function. In comparison, the procedure in this paper can be thought of as integrating out the latent variable over an infinitesimally small time block. Thus the efficiency of the Jiang and Knight (2002) procedure is greater than that achieved by the estimator in our paper (though less than that of Singleton (2001)), while at the same time being less computationally intensive than the Singleton (2001) estimator (though more than ours).

While our paper avoids the need for discretizing a continuous-time process, we do use a discrete set of moment conditions as is typical with GMM-based procedures.³ However, we should point out that a continuum of moment conditions may also be used, the advantage being that an estimator based on a continuum of moment conditions can achieve the Cramer–Rao lower bound achievable by maximum likelihood estimation. This approach has been pointed out in the context of discrete-time processes by Carrasco and Florens (2000). Subsequently, Carrasco et al. (2001) has integrated the continuum of moment conditions approach in Carrasco and Florens (2000) with

² Some other applications of this technique include Chacko (1999) and Chacko and Viceira (1999). Chacko (1999) uses this approach to estimate continuous-time models of the term structure of interest rates. Chacko and Viceira (1999) use this technique to estimate a model of stochastic precision (the reciprocal of volatility) for stock returns.

³ We thank Eric Ghysels for pointing this out and directing us to the literature utilizing a continuum of moment conditions with GMM.

the characteristic function-based estimation of diffusions employed in our paper, as well as Singleton (2001) and Jiang and Knight (2002), to produce a characteristic function-based estimation approach for diffusion processes that utilizes a continuum of moment conditions. Carrasco et al. (2001) show through Monte Carlo simulations that their integrated approach results in finite sample performance that is comparable to maximum likelihood estimation for diffusion processes.

The organization of the paper is as follows. Section 2 shows how to derive the partial differential equation (PDE) governing the characteristic function of a general continuous-time process. Section 3 outlines the direct estimation procedure utilizing the characteristic function. Section 4 discusses how to design Spectral GMM estimators to attain the efficiency of a minimum-variance estimator. Section 5, the heart of the paper, demonstrates the versatility and ease of the procedure with several applications relevant to finance. Finally, Section 6 offers concluding comments and directions for further research.

2. Characteristic functions of continuous-time stochastic processes

We start by assuming that a state variable, $X_t \in \mathbb{R}$ follows a jump–diffusion process adapted to some augmented filtration $(\mathcal{F}_t)_{t \geq 0}$ in a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. Let \mathcal{R} denote the range of this state variable. The filtration is generated by a Wiener process, W_t , and a jump (or Poisson) process, $N_t(\lambda)$. The jump process $N_t(\lambda)$ takes on a value of one when a jump occurs and it is zero otherwise. $N_t(\lambda)$ is assumed to have a constant jump frequency λ .⁴ The dynamics of the state variable is given by

$$dX_t = \mu(X_t; \boldsymbol{\theta}) dt + \sigma(X_t; \boldsymbol{\theta}) dW_t + J_t \Gamma(X_t; \boldsymbol{\theta}) dN_t(\lambda), \quad (1)$$

where $\boldsymbol{\theta}$ is a k -dimensional vector of parameters that determine the probability distribution of X_t , and $\mu(X_t; \boldsymbol{\theta})$ and $\sigma(X_t; \boldsymbol{\theta})$ represent the drift and diffusion, respectively, of the stochastic differential equation. The product $J_t \Gamma(X_t; \boldsymbol{\theta})$ represents the jump magnitude. When a jump occurs, a draw takes place from a distribution function that determines the value of J . This value is then multiplied by $\Gamma(X_t; \boldsymbol{\theta})$ to determine the magnitude of the jump.

The conditional characteristic function for X_t is defined as

$$\begin{aligned} \phi(\omega, \tau; \boldsymbol{\theta}, X_t) &= E[\exp(i\omega X_{t+\tau}) | X_t] \\ &= E[\cos(\omega X_{t+\tau}) | X_t] + iE[\sin(\omega X_{t+\tau}) | X_t], \end{aligned} \quad (2)$$

where the second line is just the Euler expansion of the exponential of a complex variable, $\tau > 0$, $i = \sqrt{-1}$ and ω represents a real-valued dummy variable. Therefore, the conditional characteristic function is simply the conditional expectation at time t of the exponentiated state variable τ periods ahead.

⁴ We use a univariate setting here for the sake of ease of exposition. However, all the results in the paper, unless otherwise specified, extend trivially to the multivariate case with multiple Wiener processes and Poisson processes. It is also easy to extend the results to include a stochastic jump frequency. In Section 5, we consider the bivariate setting of a stochastic volatility model to demonstrate that the univariate results do in fact extend easily.

The characteristic function $\phi(\omega, \tau; \boldsymbol{\theta}, X_t)$ is the integral of a complex-valued random variable. It can be shown that this integral is always finite, with $\phi(0, \tau; \boldsymbol{\theta}, X_t) = 1$ and $|\phi(\omega, \tau; \boldsymbol{\theta}, X_t)| \leq 1$ for all ω (Grimmett and Stirzaker, 1992, Theorem 5.7.3). Another important property of the characteristic function is uniqueness. If two stochastic processes have the same characteristic function, then they have the same probability distribution. Finally, we can compute all non-central moments for $X_{t+\tau}$ from the characteristic function by the formula

$$E[(X_{t+\tau})^n | \mathcal{F}_t] = \frac{1}{i^n} \left. \frac{d^n}{d\omega^n} \phi(\omega, \tau; \boldsymbol{\theta}, X_t) \right|_{\omega=0}. \quad (3)$$

Our procedure essentially uses the characteristic functions of continuous-time processes to derive moment conditions, not on the real plane as indicated by (3), but instead on the complex plane. For this procedure to be useful it needs to be shown that the conditional characteristic function can be derived in closed-form for some class of continuous-time processes. Papers by Bakshi and Madan (2000), Duffie et al. (2000), and Chacko and Das (2002) have shown that conditional characteristic functions can be derived in closed form for stochastic processes that are *exponential affine*, even though the corresponding density functions are almost always unknown.⁵ The class of exponential affine processes encompass most of the stochastic processes used currently in continuous-time finance.

The most widely used method for deriving characteristic functions for a continuous-time process is to solve the associated Kolmogorov backward equation (KBE) for the process.⁶ The KBE for the process in (1) is given by

$$\mathcal{D}\phi(\omega, \tau; \boldsymbol{\theta}, X_t) = 0, \quad (4)$$

where \mathcal{D} represents the infinitesimal generator for the process. In general, this equation is a partial differential-difference equation (PDDE) which can be solved using the boundary condition for a characteristic function:

$$\phi(\omega, 0; \boldsymbol{\theta}, X_t) = \exp(i\omega X_t) \quad (5)$$

Examples of deriving the characteristic function by solving the KBE for a process are given throughout the paper. We will focus on the class of exponential affine processes; however, it should be noted that the estimation approach via the empirical characteristic function is feasible regardless of whether the characteristic function is exponential-affine in form or not.

3. Spectral GMM estimation procedure

We have pointed out in Section 2 that unfortunately there are only a few continuous-time processes for which the conditional density function is known in closed form.

⁵ See Duffie and Kan (1993) for a definition of exponential-affine processes.

⁶ See Lo (1988) for a simple description of the KBE for a process or Karatzas and Shreve (1988) for a more rigorous treatment. Heston (1993) was the first to utilize the KBE to solve for the characteristic function of a continuous-time process in a finance setting.

Since the conditional (and unconditional) densities for most processes are unknown, it is impossible to implement a direct maximum likelihood estimation procedure and we need to resort to indirect estimation procedures such as efficient method of moments (EMM) or SMM. Section 2 also shows that for a large class of these processes the conditional characteristic function is known. Hence it would be possible to implement maximum likelihood estimation by integrating the conditional characteristic function to obtain the conditional density function. This indirect procedure, while theoretically possible, can be very expensive computationally, especially when both the dimensionality of the state vector and the sample size are large.

In this section we show that we can still use the conditional characteristic function to carry out consistent estimation of the parameter vector θ using standard GMM procedures. To see this, first note that the definition of the conditional characteristic function implies

$$E[\exp(i\omega X_{t+\tau}) - \phi(\omega, \tau; \theta, X_t) | X_t] = 0, \tag{6}$$

for all $\omega \in \mathbb{R}$. Eq. (6) defines an (infinite) set of complex-valued moment conditions. We can use the Euler expansion of the exponential function of a complex variable to transform each one of these complex-valued moment conditions into the following pair of real-valued moment conditions:

$$E[\text{Re}(\exp(i\omega X_{t+\tau}) - \phi(\omega, \tau; \theta, X_t)) | X_t] = 0,$$

$$E[\text{Im}(\exp(i\omega X_{t+\tau}) - \phi(\omega, \tau; \theta, X_t)) | X_t] = 0,$$

where $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ are real-valued operators that extract the real part and the imaginary part of a complex number. For the pair of moment conditions above, these operators give

$$\text{Re}(\exp(i\omega X_{t+\tau}) - \phi(\omega, \tau; \theta, X_t)) = \cos(i\omega X_{t+\tau}) - \text{Re}(\phi(\omega, \tau; \theta, X_t)),$$

$$\text{Im}(\exp(i\omega X_{t+\tau}) - \phi(\omega, \tau; \theta, X_t)) = \sin(i\omega X_{t+\tau}) - \text{Im}(\phi(\omega, \tau; \theta, X_t)).$$

More generally, if there is a set of (real- or complex-valued) instruments available, we have the following complex-valued unconditional moments:

$$E[\mathbf{h}(X, t) \otimes \varepsilon(\theta, \omega; t)] = \mathbf{0}, \tag{7}$$

where $\varepsilon(\theta, \omega; t) = \exp(i\omega X_{t+\tau}) - \phi(\omega, \tau; \theta, X_t)$, $\mathbf{h}(X, t) = (h_1(X, t), h_2(X, t), \dots, h_r(X, t))'$ is an r -dimensional vector of instruments orthogonal to $\varepsilon(\theta, \omega; t)$ and $\omega \in \mathbb{R}$. Once again, each of these complex-valued moment conditions imply a set of pairs of real-valued moment conditions

$$E[\text{Re}(\mathbf{h}(X, t) \otimes \varepsilon(\theta, \omega; t))] = \mathbf{0},$$

$$E[\text{Im}(\mathbf{h}(X, t) \otimes \varepsilon(\theta, \omega; t))] = \mathbf{0}.$$

Choosing a fixed grid for $\omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ (for example, $\omega = \{1, 2, \dots, n\}$),⁷ we can stack the pairs of real-valued moment conditions as

$$E[\mathbf{G}(\boldsymbol{\theta}; X, t)] = \mathbf{0}, \tag{8}$$

where $\mathbf{G}(\boldsymbol{\theta}; X, t)$ is a $(2rn \times 1)$ vector of moment conditions,

$$\mathbf{G}(\boldsymbol{\theta}; P, t) = \begin{bmatrix} \text{Re}(\mathbf{h}(X, t) \otimes \boldsymbol{\varepsilon}(\boldsymbol{\theta}; t)) \\ \text{Im}(\mathbf{h}(X, t) \otimes \boldsymbol{\varepsilon}(\boldsymbol{\theta}; t)) \end{bmatrix},$$

and $\boldsymbol{\varepsilon}(\boldsymbol{\theta}; t) = (\boldsymbol{\varepsilon}(\boldsymbol{\theta}, \omega_1; t), \boldsymbol{\varepsilon}(\boldsymbol{\theta}, \omega_2; t), \dots, \boldsymbol{\varepsilon}(\boldsymbol{\theta}, \omega_n; t))'$ is an n -dimensional vector of error terms orthogonal to $\mathbf{h}(X, t)$.

Thus we have transformed the set of complex-valued moment conditions given in (7) into the set of real-valued moment conditions given in (6). This transformation allows us to treat the characteristic-based estimation problem as a standard GMM estimation problem as follows. Given a sample of the state variable X_t observed at discrete time intervals, $t = \{t_1, t_2, \dots, t_T\}$, we can construct a sample counterpart of the expectation on the left-hand-side of (8) as

$$\mathbf{g}(\boldsymbol{\theta}; X, T) = \frac{1}{T} \sum_{i=1}^T \mathbf{G}(\boldsymbol{\theta}; X, t_i).$$

We define the Spectral GMM (SGMM) estimator of $\boldsymbol{\theta}$ as the solution to

$$\hat{\boldsymbol{\theta}}_{SGMM} = \arg \min_{\{\boldsymbol{\theta}\}} \mathbf{g}(\boldsymbol{\theta}; X, T)' \mathbf{W}(\boldsymbol{\theta}; X, T) \mathbf{g}(\boldsymbol{\theta}; X, T), \tag{9}$$

where $\mathbf{W}(\boldsymbol{\theta}; X, T)$ is a positive-definite, symmetric weighting matrix.

Therefore, the SGMM estimator of $\boldsymbol{\theta}$ will inherit the optimality properties of GMM estimators, provided that the usual regularity conditions hold (Hansen, 1982). Thus, the asymptotic variance of the SGMM estimator $\hat{\boldsymbol{\theta}}_{SGMM}$ is minimized when we choose the following (optimal) weighting matrix:

$$\mathbf{W}^*(\boldsymbol{\theta}; X, T) = \mathbf{S}^{-1},$$

where $\mathbf{S} = \lim_{T \rightarrow \infty} TE[\mathbf{g}(\boldsymbol{\theta}; X, T)\mathbf{g}(\boldsymbol{\theta}; X, T)']$. In practice, we can replace $\mathbf{W}^*(\boldsymbol{\theta}; X, T)$ with any consistent estimate. For example, if $\mathbf{g}(\boldsymbol{\theta}; X, T)$ is serially uncorrelated, a consistent estimate of \mathbf{W}^* is given by the inverse of

$$\hat{\mathbf{S}} = \frac{1}{T} \sum_{i=1}^T \mathbf{G}(\hat{\boldsymbol{\theta}}; X, t_i)\mathbf{G}(\hat{\boldsymbol{\theta}}; X, t_i)',$$

where $\hat{\boldsymbol{\theta}}$ is any consistent estimator of $\boldsymbol{\theta}$. If the vector $\mathbf{G}(\hat{\boldsymbol{\theta}}; X, t_i)$ is autocorrelated, we can use a Newey-West (1987) estimate of \mathbf{S} or any other autocorrelation and heteroskedasticity consistent estimate.⁸

⁷ In this paper, we typically use $\omega = 1, 2, \dots$. This choice is arbitrary, but choosing these values carefully can lead to more efficient estimators. See Singleton (2001) for a detailed discussion of this. Furthermore, Carrasco et al. (2001) show how to use a continuum of values for ω , which in theory can allow one to achieve the efficiency of maximum likelihood estimation.

⁸ Note that if P_t is Markov, and the instruments h depend only on P_t then $\{G_t\}$ are serially uncorrelated.

Given any sequence of optimal weighting matrices, the SGMM estimator has the following properties (see Proposition 14.1 in Hamilton, 1994):

1. Consistency: $\hat{\boldsymbol{\theta}}_{SGMM} \xrightarrow{p} \boldsymbol{\theta}$.
2. Asymptotic normality: $\sqrt{T}(\hat{\boldsymbol{\theta}}_{SGMM} - \boldsymbol{\theta}) \xrightarrow{p} N(\mathbf{0}, \mathbf{V})$, where $\mathbf{V}^{-1} = \mathbf{D}\mathbf{W}^*\mathbf{D}'$, and

$$\mathbf{D}' = \text{plim} \left\{ \frac{\hat{\partial} \mathbf{g}(\hat{\boldsymbol{\theta}}_{SGMM}; X, t_i)}{\hat{\partial} \boldsymbol{\theta}'} \right\}.$$

4. Spectral GMM and ML estimation

Maximum likelihood (ML) estimators are globally efficient and unbiased. Whenever there exists an unbiased estimator whose variance attains the Cramér–Rao (CR) bound, the ML estimator coincides with this (Silvey, 1975). GMM estimators are not generally equivalent to ML estimators and hence they do not share with them this desirable optimality property. However, with the use of an appropriate weighting function and a continuum of moment conditions, GMM estimators can be made equivalent to ML and attain the CR bound. We demonstrate that in this section using arguments adopted from Feuerverger and McDunnough (1981a,b). We keep the discussion brief here and refer the reader to Feuerverger and McDunnough (1981a,b) and Singleton (2001) for further details regarding the optimal weighting function, and Feuerverger (1990) and Singleton (2001) regarding the issue of the appropriate set of moment conditions.

To see this, note that the gradient of the likelihood function verifies

$$\begin{aligned} E \left[\frac{\partial \log f(X_{t+\tau}, \tau | \boldsymbol{\theta}, X_t)}{\partial \boldsymbol{\theta}} \mid \mathcal{F}_t \right] &= \int_{-\infty}^{\infty} \frac{\partial \log f(X_{t+\tau}, \tau | \boldsymbol{\theta}, X_t)}{\partial \boldsymbol{\theta}} f(X_{t+\tau}, \tau | \boldsymbol{\theta}, X_t) dX_{t+\tau} \\ &= 0. \end{aligned} \tag{10}$$

But substituting the conditional characteristic function for $f(X_{t+\tau}, \tau | \boldsymbol{\theta}, X_t)$ in (10) and reordering terms we obtain

$$\begin{aligned} 0 &= \int_{-\infty}^{\infty} \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\partial \log f(X_{t+\tau}, \tau | \boldsymbol{\theta}, X_t)}{\partial \boldsymbol{\theta}} \exp(-i\omega X_{t+\tau}) dX_{t+\tau} \right] \phi(\omega, \tau; \boldsymbol{\theta}, X_t) d\omega \\ &\equiv \int_{-\infty}^{\infty} h(\omega, X_t) \phi(\omega, \tau; \boldsymbol{\theta}, X_t) d\omega, \end{aligned}$$

which implies immediately the following moment condition:

$$E \left[\int_{-\infty}^{\infty} h(\omega, X_t) [\exp(i\omega X_{t+\tau}) - \phi(\omega, \tau; \boldsymbol{\theta}, X_t)] d\omega \right] = 0. \tag{11}$$

This transformation shows that moment conditions (10) and (11) are equivalent. This implies in turn that a GMM estimator based on the moment condition (10) will be equivalent to a Spectral GMM estimator based on the complex-valued moment condition (11). But the GMM estimator based on the moment condition (10) is just

the ML estimator. Therefore, the Spectral GMM estimator based on (11) must also be ML and globally efficient.

Eq. (11) shows that an appropriate choice of instruments in the Spectral GMM procedure will render GMM estimates that are also globally efficient ML estimates. These instruments are in fact a continuum of instruments indexed by ω .⁹ These instruments also depend on the particular form of the density function. However, we can still substitute $h(\omega, t)$ for any consistent estimate, and the integral for a discrete partition that is fine enough, and the resulting estimates will still verify (11) asymptotically.

5. Applications to finance

In this section we present some applications relevant for finance for which estimation using maximum likelihood is difficult because an analytical expression for the density function of the stochastic process is unknown, but for which estimation via Spectral GMM is simple, because the characteristic function is known. Because the differences between the technique in this paper and those in Singleton (2001) and Jiang and Knight (2002) occur with latent variable models, our applications are focused exclusively on stochastic volatility models.

We first estimate a stochastic volatility model for stock prices. Next we estimate a pure jump–diffusion model with constant volatility, and then a combined stochastic volatility, jump–diffusion model. This combination is interesting, because it allows us to better understand the contribution of each component to explain the excess kurtosis and skewness observed in stock return data. We next show how Spectral GMM may be employed in non-affine settings in the context of a stochastic volatility model. We estimate a model where the diffusion term on the variance process is proportional to variance raised to an arbitrary power, similar to the Chan et al. (1992) specification for interest rates. Finally, we also use Spectral GMM to show some evidence of the presence of high and low frequency components in stock return volatility.

A number of papers have estimated models of the types used in this section. We pick a representative sample of the most recent papers and we compare the results obtained in our paper against these. The single-factor, square root stochastic volatility model (with no jumps) has been estimated by Jiang and Knight (2002) and an earlier (working paper) version of Singleton (2001)¹⁰ using empirical characteristic function estimation. Because these two papers are the closest in methodology to ours we include these in our comparison set. The results of this section will also be compared with other recent papers estimating similar models (though with differing empirical

⁹ It should be noted, however, that in practice a continuum of moment conditions is difficult to implement. The decision of which moments to choose is a difficult issue. Singleton (2001) contains a discussion regarding this.

¹⁰ While the published version of Singleton (2001) does not contain these empirical results, we nevertheless thought it would be valuable to compare our empirical results to those that would be obtained from Singleton's (2001) procedure with the same data set. Hence, we used the empirical results that were contained in earlier versions of the Singleton paper. The empirical results that are contained in these earlier versions of Singleton (2001) can be obtained from Ken Singleton, or from the authors of this paper.

methods), including Chernov and Ghysels (2000), Chernov et al. (2000) (CGGT), Andersen et al. (2002), Chernov et al. (2002), and Pan (2002).

5.1. Stochastic volatility

A great deal of research in finance, beginning with Black (1976), has looked at the implications of time-varying volatility for asset prices. This research includes stochastic volatility models, such as those in Wiggins (1987), Hull and White (1987), Melino and Turnbull (1990), Stein and Stein (1991), and Amin and Ng (1993), as well as GARCH models, such as Engle (1982), Bollerslev (1986), Nelson (1989), and Hentschel (1995).¹¹ In this section, we estimate a model of stock price dynamics where the instantaneous volatility of the stock price is stochastic. We augment the basic geometric Brownian motion model with a square-root model for volatility as follows:

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{v_t} dW_S, \quad (12)$$

$$dv_t = \kappa(\theta - v_t) dt + \sigma\sqrt{v_t} dW_v, \quad (13)$$

where v_t represents the instantaneous variance of the stock price. The parameters μ , κ , θ , and σ are all constants. The instantaneous correlation between W_S and W_v is a constant ρ .

In this model for stock return dynamics, volatility is an unobservable stochastic variable. To estimate the parameters of this process by traditional maximum likelihood methods we would need first to obtain the density function for the stock price conditional on the current stock price and volatility. Next we would need to integrate volatility out of the density function to obtain the density function for the stock price conditional only on the current stock price. Unfortunately, there is no known analytical expression for the conditional density function. This fact makes impossible direct maximum likelihood estimation, and it has led to an explosion of research on numerical methods that can be helpful to attack this problem.¹² However, all of the methods developed so far have been computationally intensive because they have to deal simultaneously with solving numerically for the conditional density and integrating volatility out of this density. By contrast, we can easily estimate this stochastic volatility model using Spectral GMM, because we can derive a closed-form expression for the conditional characteristic function of this process.

¹¹ Nelson (1990) shows that GARCH processes converge in distribution to diffusion processes as the time interval shrinks. Bollerslev et al. (1992) provide a comprehensive review of the use of GARCH models in Finance.

¹² Common techniques used have included GMM (and the EMM procedure of Gallant and Tauchen (1996)), Kalman filtering, simulated maximum likelihood, and Bayesian estimation. The choice between these usually becomes a tradeoff between accuracy and computation time. See Melino and Turnbull (1990), Gallant et al. (1997), Harvey et al. (1994), Danielsson (1994), and Jacquier et al. (1994) for examples of these estimation methods in the context of stochastic volatility models. See Ghysels et al. (1996) for a literature review on estimation methods associated with stochastic volatility models.

To derive the conditional characteristic function for (12), we first transform (12) so that it satisfies the conditions in Duffie et al. (2000) and Chacko and Das (2002) for an exponential-affine solution. The transformed model is given by

$$\begin{aligned}
 d \log S_t &= \left(\mu - \frac{1}{2} v_t \right) dt + \sqrt{v_t} dW_S, \\
 dv_t &= \kappa(\theta - v_t) dt + \sigma \sqrt{v_t} dW_v.
 \end{aligned}
 \tag{14}$$

The conditional characteristic function, $\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t, v_t)$,¹³ satisfies the following PDE:

$$\mathcal{D}\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t, v_t) = 0,
 \tag{15}$$

where

$$\begin{aligned}
 \mathcal{D}\phi &= \frac{1}{2} v_t \frac{\partial^2 \phi}{\partial \log S_t^2} + \rho \sigma v_t \frac{\partial^2 \phi}{\partial \log S_t \partial v_t} + \frac{1}{2} \sigma^2 v_t \frac{\partial^2 \phi}{\partial v_t^2} + \left(\mu - \frac{1}{2} v_t \right) \frac{\partial \phi}{\partial \log S_t} \\
 &\quad + \kappa(\theta - v_t) \frac{\partial \phi}{\partial v_t} - \frac{\partial \phi}{\partial \tau}.
 \end{aligned}$$

The boundary condition for (15) is given by $\phi(\omega, 0; \boldsymbol{\theta}, \log S_T, v_T) = \exp(i\omega \log S_T)$. For details on the derivation of this PDE and others in this paper, see Duffie et al. (2000) or Chacko and Das (2002).

Eq. (15) has an exact solution given by

$$\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t, v_t) = \exp[i\omega \log S_t + A(\omega, \tau; \boldsymbol{\theta})v_t + B(\omega, \tau; \boldsymbol{\theta})],
 \tag{16}$$

where

$$\begin{aligned}
 A(\omega, \tau; \boldsymbol{\theta}) &= \frac{2}{\sigma^2} \left[\frac{u_1 u_2 e^{u_1 \tau} - u_1 u_2 e^{u_2 \tau}}{u_1 e^{u_2 \tau} - u_2 e^{u_1 \tau}} \right], \\
 B(\omega, \tau; \boldsymbol{\theta}) &= \mu i \omega \tau + \frac{2\kappa\theta}{\sigma^2} \log \left[\frac{u_2 - u_1}{u_2 e^{u_1 \tau} - u_1 e^{u_2 \tau}} \right], \\
 u_1 &= \frac{1}{2} \left[\rho \sigma i \omega - \kappa + \sqrt{(\rho \sigma i \omega - \kappa)^2 - \sigma^2 i \omega (i \omega - 1)} \right], \\
 u_2 &= \rho \frac{1}{2} \left[\sigma i \omega - \kappa - \sqrt{(\rho \sigma i \omega - \kappa)^2 - \sigma^2 i \omega (i \omega - 1)} \right].
 \end{aligned}$$

Eq. (16) is the characteristic function of the stock price conditional on the current stock price and volatility, which is unobservable. To estimate the parameters of the

¹³ We are using a shortcut notation here. To be precise, we should first define the joint characteristic function of the stock price and volatility as

$$\phi(\omega_1, \omega_2, \tau; \boldsymbol{\theta}, \log S_t, v_t) = E[\exp(i\omega_1 \log S_{t+\tau} + i\omega_2 v_{t+\tau}) | \log S_t, v_t]$$

The characteristic function notation that we use throughout the paper would then be defined as that with ω_2 set equal to 0 always, i.e.,

$$\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t) \equiv \phi(\omega_1, 0, \tau; \boldsymbol{\theta}, \log S_t, v_t),$$

stochastic volatility model we need first to integrate out the unobservable variable out of this function:

$$\begin{aligned}\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t) &= \int_0^\infty \phi(\omega, \tau; \boldsymbol{\theta}, \log S_t, v_t) f(v_t) dv_t \\ &= \exp[i\omega \log S_t + B(\omega, \tau; \boldsymbol{\theta})] \int_0^\infty \exp[A(\omega, \tau; \boldsymbol{\theta})v_t] f(v_t) dv_t.\end{aligned}$$

where $f(v_t)$ represents the unconditional density of v_t . The integral $\int_0^\infty \exp[A(\omega, \tau; \boldsymbol{\theta})v_t] f(v_t) dv_t$ can simply be thought of as a version of the unconditional characteristic function of v_t , where instead of $i\omega$ we have a slightly more complicated expression in $A(\omega, \tau; \boldsymbol{\theta})$.¹⁴ In general, any affine latent variable model can be estimated by following this method of integrating the latent variable out of the characteristic function.

It is important to note that what we are calling the conditional characteristic function in this particular example, as well as in subsequent examples involving latent variables, is not entirely conditional in the sense that we do not condition on all of the information available in the previous time period. Specifically, we do not condition on the entire path of the stock price, but rather on the level of the stock price in the previous period. However, due to the correlation between stock returns and volatility, the level of the stock price alone in the previous period does not contain all the available information about the conditional distribution of the following period's stock price, i.e., the stock price alone is non-Markov.¹⁵ Therefore, the entire path that the stock price takes to some time t contains information about the level of volatility at time t . By not conditioning on this information we lose efficiency, but the trade-off is that we gain immensely in terms of computational speed. Singleton (2001) and Jiang and Knight (2002) achieve greater efficiency at the cost of greater computation time by constructing estimators that utilize more information along the historical path of stock prices than utilized here.¹⁶ Finally, it is important to remember that the loss of efficiency relative to Singleton (2001) and Jiang and Knight (2002) occurs only with latent variable models and not generally with observable multi-factor models or jump–diffusion based models.

Integrating the variance v_t out of the characteristic function leads to the following expression for the characteristic function conditional only on the stock price:

$$\begin{aligned}\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t) &= \exp \left[i\omega \log S_t + B(\omega, \tau; \boldsymbol{\theta}) \right. \\ &\quad \left. + \frac{2\kappa\theta}{\sigma^2} \log \left(\frac{2\kappa}{2\kappa - \sigma^2 A(\omega, \tau; \boldsymbol{\theta})} \right) \right].\end{aligned}\tag{17}$$

¹⁴ The conditional moment generating function, $\phi_t(\tau)$, for any random variable, x_t , satisfies the same PDE as the characteristic function, but subject to a different boundary condition given by $\phi_t(0) = \exp[\omega x_t]$. The unconditional moment generating function obtains as $\lim_{\tau \rightarrow \infty} \phi_t(\tau)$.

¹⁵ Note however that the stock price and volatility form a Markov system, but the problem in these types of latent variable models is that volatility is unobservable. This creates the need to calculate the density of the stock price conditional only on past stock prices.

¹⁶ Singleton (2001) uses a simulated method of moments procedure utilizing a fully conditional characteristic function (using the entire path of stock prices), while Jiang and Knight (2002) proposes an estimator that utilizes part (ranging from two to six prior observations) of the historical path of stock prices.

Table 1

Parameter estimates for the stochastic volatility model

This table presents parameter estimates (standard errors) for the stochastic volatility model

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{v_t} dW_{s,t},$$

$$dv_t = \kappa(\theta - v_t) dt + \sigma\sqrt{v_t} dW_{v,t},$$

where

$$\text{Corr}(dW_{s,t}, dW_{v,t}) = \rho.$$

The column “Daily1” represents parameter estimates using daily stock price data from 1990–1999, the column “Daily2” represents parameter estimates using daily stock price data for 1980–2000, the column “Weekly” represents estimates using weekly data for 1962–2000, while the column “Monthly” represents estimates using monthly data for 1926–1997.

Parameter	Daily1 data	Daily2 data	Weekly data	Monthly data
μ	0.1406 (0.0573)	0.1121 (0.0379)	0.1028 (0.0239)	0.1204 (0.0257)
κ	16.6997 (4.5364)	14.2819 (3.2753)	2.5371 (2.1627)	0.6309 (0.3528)
θ	0.0255 (0.0039)	0.0330 (0.0021)	0.0247 (0.0028)	0.0371 (0.0051)
σ	4.7148 (0.5860)	5.1925 (0.4722)	1.0370 (0.1092)	0.2611 (0.0340)
ρ	-0.4135 (0.2269)	-0.6289 (0.2097)	-0.5163 (0.2625)	-0.6714 (0.4180)

With the unobservable state variable integrated out, we can now apply a Spectral GMM procedure to estimate the parameters of the process (12)–(13) using stock price data. Note that we can compute the n -th conditional moment of $\exp(\log S_t)$ by simple substitution of $\omega = n$ into (17). We can then use these moments in a conditional GMM procedure.

Table 1 reports Spectral GMM estimates of the process (12)–(13) for the CRSP value-weighted portfolio measured at three different frequencies, monthly (from January 1926 to December 2000) and weekly (from the first week of 1962 to the last week of 2000). In addition we provide estimates using daily data on the S&P 500 index from January 1980 to December 2000 as well as from January 1990 to December 1999.¹⁷ The daily data is provided primarily to facilitate comparison with other papers that have estimated stochastic volatility models. It is important to note that while data from the S&P 500 index with a daily sampling frequency is used in our paper as well as the papers we compare amongst, all of the papers use differing time periods for estimation, so the results are not entirely comparable. However, because the time

¹⁷ We use two sets of daily data for this subsection only. This is because Jiang and Knight (2002) and Singleton (2001), the papers that are closest in methodology to ours, use data from 1990 to 1999, while the other papers we compare with use various other time periods. By providing estimates for 1990 to 1999, the comparisons with Jiang and Knight (2002) and Singleton (2001) will be the most direct and valid of the comparisons we do.

periods do overlap among all of the papers, we use the comparisons merely to get a “sense” of how estimation results vary across papers.

We estimate the stochastic volatility model using the first five spectral moments, i.e., for $\omega = 1, 2, \dots, 5$. For ease of comparing estimates produced from different data frequencies, we also set τ such that the resulting parameter estimates are annualized—for example, with monthly data, we set $\tau = 1/12$ and with weekly data we set $\tau = 1/52$. With annualized parameter estimates the reader does not have to concern himself with the frequency of the data in comparing parameter estimates; estimates produced from different data sets and different sampling frequencies can be directly compared with each other. Finally, $\mathbf{h}(\cdot, \cdot)$ is a vector of ones.

The estimation results show that stochastic volatility is clearly an important factor in stock price dynamics: The parameter σ , which premultiplies shocks to volatility, is strongly statistically significant both in the monthly and in the weekly data. This parameter is important to capture excess kurtosis in the data, and its strong significance indicates the importance of this feature in stock price dynamics. Furthermore, the estimate of σ from weekly returns is much larger than the estimate from monthly returns, while the estimate from daily returns is higher still, implying that the distribution of daily and weekly returns has much fatter tails than the distribution of monthly returns. The unconditional mean (θ) of stock volatility is also strongly significant in all three datasets, and the estimate from weekly returns is lower than the estimate from monthly returns.

The correlation coefficient, ρ , captures skewness in the distribution of stock returns. The negative sign indicates the presence of negative skewness in stock returns. Interestingly, the correlation is larger in magnitude with monthly data than weekly data. This suggests greater skewness in monthly returns than in weekly returns. However, this may also occur simply because the monthly data comprises the 1926–1940 period, in which there are many large negative returns.

The estimate for κ , the rate of mean reversion in volatility, is much lower for monthly returns than for weekly returns, while the daily return estimate is even higher. This has strong implications for the persistence of shocks to volatility at different frequencies. The estimate of κ for weekly returns implies a half-life of a shock to volatility of 3.3 months, while the estimate with daily data implies a half-life of 0.6 months.¹⁸ By contrast, there appears to be much more persistence when we measure returns at a monthly frequency. In that case the estimate of κ implies a the half-life of a shock of 13.1 months. This suggests the presence of both short and long-run components in volatility. Section 5.5 explores this issue in detail, including the pattern of rising estimates for κ obtained as the data frequency increases. The rising values for κ with

¹⁸ The half-life of a process is simple to compute; if we take a stochastic process at its unconditional mean and shock the process, then the expected time it would take the process to return half way to its unconditional mean is its half-life. For a square root process, the conditional mean of the process is given by the expression

$$E(v_t | v_0) = v_0 e^{-\kappa t} + \theta(1 - e^{-\kappa t}).$$

So, if we shock the process so that $v_0 = 2\theta$, then the amount of time t it takes for v to get to a value of 1.5θ is $1/\kappa \log 2$.

data frequency also explain why we observe σ increasing with data frequency. As κ increases, volatility becomes less persistent; thus σ , the volatility of volatility, needs to rise to help explain the variation of volatility observed in the data. In other words, the drift of the volatility process explains less and less of the variation in volatility, causing the data to load more heavily on the diffusion term in the volatility process.

In comparing our results with daily S&P 500 data (between 1990 and 1999) against those found in Singleton (2001) and Jiang and Knight (2002), we find that our results are in line with those found in both of these papers. The only area where our results differ significantly is in the estimate of κ . Singleton (2001) and Jiang and Knight (2002) obtain half-life estimates of between 2 and 3 days. Our estimate, on the other hand, implies a half-life for our process of roughly 10 days. The explanation for this difference is likely the fact that because Singleton's (2001) and Jiang and Knight's (2002) procedures utilize more time-series information than ours, they should be able to capture the dynamic properties of the stochastic volatility process better. This seems to be exhibiting itself mostly in the estimate of the mean reversion parameter. This has important economic implications; for example, Chacko and Viceira (1999) show that portfolio choice decisions can be very sensitive to the mean reversion parameter. The other difference to note between our results and those of Singleton (2001) and Jiang and Knight (2002) is that our standard errors are higher: this is to be expected as our integration procedure results in a loss of information, and therefore, efficiency. It is difficult to make more general conclusions about our approach versus those in Singleton (2001) and Jiang and Knight (2002) without the use of carefully done simulation studies; however, these studies are beyond the scope of this paper and are currently being undertaken as part of another project.

The papers utilizing efficient method of moments (EMM) all obtain significantly lower rates of mean reversion than us. In Andersen et al. (2002), for example, the estimate of the half-life for the same stochastic volatility model ranges from 43 to 54 days, depending on the length of the time series. Meanwhile, Chernov and Ghysels (2000) obtains a half-life estimate of 188 days. Similar results are found in Chernov et al. (2002), though Chernov et al. (2002) seems to obtain slightly lower values.¹⁹ These results, while indicating a disparity between the estimation techniques, can be explained intuitively. As discussed above, the higher the estimate of the mean reversion parameter, the higher the estimate of σ that is obtained because the decrease in volatility persistence needs to be offset by an increase in the volatility of volatility in order to explain the conditional volatility observed in the data. A more formal way of seeing this is to look at the expression for conditional volatility of volatility:

$$\text{Var}[v_T|v_t] = \frac{\sigma^2}{2\kappa} (1 - e^{-2\kappa(T-t)})$$

In this expression, for a given value of conditional variance, as the estimate for κ rises, the higher the value of σ must be in order to match the given conditional variance. While we obtain an estimate for κ of 14.3 (using data from 1980 to 2000), Chernov

¹⁹ These models utilize stochastic drift terms as well, thus affecting the estimates of volatility. Effectively, some of the variation in stock price that is due to volatility in our model and others is explained by the time-varying drift, thus dampening the volatility process.

and Ghysels (2000) estimate a value of 0.9 (using data from 1985 to 1993). Therefore, our estimate for σ must be higher than theirs to explain the variance of volatility. This is why we obtain an estimate of 5.2 while theirs is 0.06. There remains a large disparity between the two estimates, but this is likely due to difference in the time period covered by the data sets. Pan (2002), utilizing an alternative estimation procedure, obtains a half-life estimate that is in between ours and that of Chernov and Ghysels (2000); however her volatility of volatility estimate is also in between ours and Chernov and Ghysels (2000), as predicted.

In conclusion, the various papers estimating stochastic volatility models all seem to agree on most parameter estimates. However, the estimates for the rate of mean reversion, κ , and the volatility of volatility, σ , seem to be difficult to pin down independently. These two values seem to vary widely across papers, so it would seem that further work exploring how to better nail down these estimates would be useful. This is particularly so due to the economic importance of these parameters: the rate of mean reversion has been shown to be important for portfolio choice decisions, while the volatility of volatility is of importance in pricing derivative securities.

5.2. Jump–diffusion process

Jump–diffusion processes are regularly used in finance to capture discontinuous behavior in asset pricing.²⁰ Return discontinuities typically exhibit themselves in discretely-sampled data in the form of excess kurtosis. In this section we estimate via a Spectral GMM procedure the following jump–diffusion process for stock price dynamics with asymmetric upward and downward jumps:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dZ + [\exp(J_u) - 1] dN_u(\lambda_u) + [\exp(-J_d) - 1] dN_d(\lambda_d), \quad (18)$$

where μ and σ are constants, $J_u, J_d > 0$ are stochastic jump magnitudes, and $\lambda_u, \lambda_d > 0$ are constants that determine jump frequencies. Hence $[\exp(J_u) - 1] dN_u(\lambda_u)$ represents a positive jump and $[\exp(-J_d) - 1] dN_d(\lambda_d)$ represents a downward jump. Note that $J_u, J_d > 0$ implies that the stock price will remain non-negative. We assume that the jump magnitudes are determined by draws from exponential distributions, with densities

$$f(J_u) = \frac{1}{\eta_u} \exp\left(-\frac{J_u}{\eta_u}\right), \quad (19)$$

and

$$f(J_d) = \frac{1}{\eta_d} \exp\left(-\frac{J_d}{\eta_d}\right).$$

²⁰ After introduction of continuous-time financial methods by Black and Scholes (1973) and Merton (1973), Merton (1976) pioneered the use of jump processes in continuous-time finance. Recent examples include Ahn and Thompson (1988), Jorion (1988), Bates (1996), and Das and Foresi (1996).

The combination of the Normal process plus a mixed Poisson-Exponential process in the jump–diffusion model results in a conditional density function for S_t that is unknown. In addition, with discretely sampled data, it is difficult to tell which returns have a discontinuous component(s) in them and which ones do not. Therefore, estimating this process using the estimation procedures currently available is extremely difficult. Even the standard Euler discretization scheme does not work here because the jump term, which contains Poisson and Exponentially distributed components, cannot be well approximated with a Normally distributed shock.

By contrast, it is straightforward to derive the conditional characteristic function of this process. This provides a simple, consistent procedure to estimate this process via spectral GMM. To derive the conditional characteristic function, we first utilize a log transformation of (18):

$$d \log S_t = \left(\mu - \frac{1}{2} \sigma^2 \right) dt + \sigma dZ + J_u dN_u(\lambda_u) - J_d dN_d(\lambda_d).$$

Next we need to derive the conditional characteristic function, $\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t)$, for the log stock price. The characteristic function satisfies the equation

$$\mathcal{D}\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t) = 0, \tag{20}$$

where

$$\begin{aligned} \mathcal{D}\phi &= \frac{1}{2} \sigma^2 \frac{\partial^2 \phi}{\partial \log S_t^2} + \left(\mu - \frac{1}{2} \sigma^2 \right) \frac{\partial \phi}{\partial \log S_t} - \frac{\partial \phi}{\partial \tau} \\ &+ \lambda_u E_t[\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t + J_u) - \phi(\omega, \tau; \boldsymbol{\theta}, \log S_t)] \\ &+ \lambda_d E_t[\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t - J_d) - \phi(\omega, \tau; \boldsymbol{\theta}, \log S_t)]. \end{aligned}$$

The boundary condition for (20) is given by $\phi(\omega, 0; \boldsymbol{\theta}, \log S_T) = \exp(i\omega \log S_T)$.

Solving (20), we obtain the characteristic function for (18). This function is given by

$$\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t) = \exp[i\omega \log S_t + A(\omega, \tau; \boldsymbol{\theta})],$$

where

$$A(\omega, \tau; \boldsymbol{\theta}) = \frac{1}{2} \sigma^2 (i\omega)^2 \tau + \left(\mu - \frac{1}{2} \sigma^2 \right) i\omega \tau + \frac{\lambda_u \tau}{1 - i\omega \eta_u} + \frac{\lambda_d \tau}{1 - i\omega \eta_d} - (\lambda_u + \lambda_d) \tau.$$

With the characteristic function known in closed-form, we can now apply the spectral GMM procedure. We use the same data as that used for the stochastic volatility model above. We estimate the process using the first six spectral moments, i.e., for $\omega = 1, 2, \dots, 6$. For ease of comparing estimates, we also set τ such that the resulting parameter estimates are annualized. Finally, $\mathbf{h}(\cdot, \cdot)$ is a vector of ones.

Table 2

Parameter estimates for the jump–diffusion model

This table presents parameter estimates (standard errors) for the jump–diffusion model of stock prices given by

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_{s,t} + [\exp(J_u) - 1] dN_u(\lambda_u) - [\exp(-J_d) - 1] dN_d(\lambda_d),$$

where the jump magnitudes J_u and J_d are draws from exponential distributions with means η_u and η_d , respectively. The column “Daily” represents parameter estimates using daily stock price data for 1980–2000, the column “Weekly” represents estimates using weekly data for 1962–2000, while the column “Monthly” represents estimates using monthly data for 1926–1997.

Parameter	Daily data	Weekly data	Monthly data
μ	0.1926 (0.0537)	0.1831 (0.0392)	0.1448 (0.0309)
σ	0.1236 (0.0297)	0.1067 (0.0269)	0.1729 (0.0447)
η_u	0.0492 (0.0088)	0.0458 (0.0147)	0.0095 (0.0045)
λ_u	1.5390 (0.2115)	1.2103 (0.2993)	1.9529 (10.0580)
η_d	0.0476 (0.0063)	0.0432 (0.0083)	0.0177 (0.0071)
λ_d	3.6329 (0.7238)	3.0143 (1.1875)	2.4067 (1.0849)

Table 2 displays the estimation results. The jump components are all statistically significant both in the weekly data and in the monthly data. They capture both the skewness and excess kurtosis present in stock returns data. The skewness is captured by the asymmetry in upward versus downward jumps. The estimates of λ_u and λ_d , which may be interpreted as the number of positive and negative jumps per annum, respectively, imply that upward jumps occur less frequently than downward jumps. This is particularly true in the weekly and daily data. Additionally, the estimates of η_u and η_d imply that the average magnitude of a negative jump is larger than the average magnitude of a positive jump in monthly data, while they are about the same in weekly and daily data. These results indicate that the data is negatively skewed, causing the model to have a higher loading on expected downward jumps versus expected upward jumps.

The negative skewness is an important trait in returns data and is commonly observed to play a substantial role in model estimation. Pan (2002) estimates a negative mean jump size (both on an objective and risk-neutral basis), and excess kurtosis is captured by the jump frequency of roughly 27 jumps per year. The jump frequency in this paper is significantly higher than ours, but excess kurtosis is captured in her model through a higher jump variance than what we calculate. In Andersen et al. (2002), a zero-mean jump magnitude is imposed so that they are not able to capture skewness through jumps (though as discussed later, they capture skewness through the correlation term between returns and volatility), but their jump frequency estimates are of the same order of magnitude as ours. Chernov et al. (2000, 2002), like Andersen et al. (2002),

nest both stochastic volatility and jumps in their models, but they do not impose a zero-mean restriction. They too estimate a negative mean jump size and their jump frequency estimates are lower than those of Pan (2002) and very similar to ours. Thus, the negative skewness in returns is captured in these models by the jump process, though only partially. Their model specifications load on the correlation parameter as well, so time-varying volatility seems to explain some of the empirical skewness results as well.

Thus, negative skewness in stock returns exhibits itself in the form of negative correlation between stock returns and volatility with a stochastic volatility model and more downward jumps with a jump–diffusion, constant volatility model. Meanwhile, the presence of jumps of time-varying volatility is sufficient to capture the excess kurtosis in stock returns. The question as to which of these models better fits these higher moments of stock returns is addressed in the next section.

5.3. Mixed stochastic volatility, jump–diffusion model

The question of whether jump processes or stochastic volatility better describe stock price dynamics has been a long running debate in financial modelling.²¹ In this section, we nest the models of the previous two section in one model to determine what is the contribution of stochastic volatility and jump processes to stock price dynamics. We estimate the following model for stock price dynamics:

$$\begin{aligned}
 d \log S_t &= \left(\mu - \frac{1}{2} v_t \right) dt + \sqrt{v_t} dW_S + [\exp(J_u) - 1] dN_u(\lambda_u) \\
 &\quad + [\exp(-J_d) - 1] dN_d(\lambda_d), \\
 dv_t &= \kappa(\theta - v_t) dt + \sigma \sqrt{v_t} dW_v,
 \end{aligned}
 \tag{21}$$

where the parameters are defined as in the previous in the two section.

From (4), the conditional characteristic function of the process, $\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t)$, satisfies the following PDE:

$$\mathcal{D}\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t, v_t) = 0,
 \tag{22}$$

where

$$\begin{aligned}
 \mathcal{D}\phi &= \frac{1}{2} v_t \frac{\partial^2 \phi}{\partial \log S_t^2} + \rho \sigma v_t \frac{\partial^2 \phi}{\partial \log S_t \partial v_t} + \frac{1}{2} \sigma^2 v_t \frac{\partial^2 \phi}{\partial v_t^2} + \left(\mu - \frac{1}{2} v_t \right) \frac{\partial \phi}{\partial \log S_t} \\
 &\quad + \kappa(\theta - v_t) \frac{\partial \phi}{\partial v_t} - \frac{\partial \phi}{\partial \tau} + \lambda_u E_t[\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t + J_u) - \phi(\omega, \tau; \boldsymbol{\theta}, \log S_t)] \\
 &\quad + \lambda_d E_t[\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t - J_d) - \phi(\omega, \tau; \boldsymbol{\theta}, \log S_t)]
 \end{aligned}$$

The solution to (22) subject to the boundary condition $\phi(\omega, 0; \boldsymbol{\theta}, \log S_T, v_T) = \exp(i\omega \log S_T)$ gives us the conditional characteristic function for the process (21). This solution is simply a combination of the characteristic functions for the stochastic volatility

²¹ See, for example, Das and Sundaram (1999) and the citations within.

model estimated in Section 5.1 and the jump–diffusion model estimated in Section 5.2. The characteristic function is given by

$$\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t, v_t) = \exp[i\omega \log S_t + A(\omega, \tau; \boldsymbol{\theta})v_t + B(\omega, \tau; \boldsymbol{\theta})],$$

where

$$A(\omega, \tau; \boldsymbol{\theta}) = \frac{2}{\sigma^2} \left[\frac{u_1 u_2 e^{u_1 \tau} - u_1 u_2 e^{u_2 \tau}}{u_1 e^{u_2 \tau} - u_2 e^{u_1 \tau}} \right],$$

$$B(\omega, \tau; \boldsymbol{\theta}) = \frac{2\kappa\theta}{\sigma^2} \log \left[\frac{u_2 - u_1}{u_2 e^{u_1 \tau} - u_1 e^{u_2 \tau}} \right] + \frac{1}{2} \sigma^2 (i\omega)^2 \tau + \left(\mu - \frac{1}{2} \sigma^2 \right) i\omega \tau$$

$$+ \frac{\lambda_u \tau}{1 - i\omega \eta_u} + \frac{\lambda_d \tau}{1 - i\omega \eta_d} - (\lambda_u + \lambda_d) \tau,$$

$$u_1 = \rho \sigma i\omega - \kappa + \sqrt{(\rho \sigma i\omega - \kappa)^2 - \sigma^2 i\omega (i\omega - 1)},$$

$$u_2 = \rho \sigma i\omega - \kappa - \sqrt{(\rho \sigma i\omega - \kappa)^2 - \sigma^2 i\omega (i\omega - 1)}.$$

As we did with the pure stochastic volatility process, we now proceed to integrate volatility out of the conditional characteristic function, so that the resulting characteristic function is conditional only on the stock price. This gives

$$\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t) = \exp \left[i\omega \log S_t + B(\omega, \tau; \boldsymbol{\theta}) + \frac{2\kappa\theta}{\sigma^2} \log \left(\frac{2\kappa}{2\kappa - \sigma^2 A(\omega, \tau; \boldsymbol{\theta})} \right) \right].$$

We can now apply spectral GMM to estimate this model using stock price data. We do so using the same data used in the previous two section. We use the first nine spectral moments of the process. For ease of comparing estimates produced from different data frequencies, we also set τ such that the resulting parameter estimates are annualized—for example, with monthly data, we set $\tau = 1/12$ and with weekly data we set $\tau = 1/52$. Finally, $\mathbf{h}(\cdot, \cdot)$ is a vector of ones.

Table 3 reports the estimation results. It is immediately apparent from this table that both stochastic volatility and jumps are needed to capture stock return dynamics. The coefficients for many of the jump components as well as σ , the volatility coefficient of stock variance, are statistically significant. However, it is interesting to note that the estimate of the upward jump frequency, λ_u , drops considerably from the jump–diffusion only (JDO) model of the previous section. The estimates of the upward jump frequency from the JDO model reported in Table 2 imply positive jumps occurring on average one to two times a year. In the stochastic volatility, jump–diffusion (SVJD) model this frequency drops to once every 3 years with the daily data, once every 5 years according to the weekly data, and only once every 35 years in the monthly data. Thus, it would seem that stochastic volatility is more important than jumps in explaining infrequent large positive stock returns.

By contrast, the estimate of the frequency of large negative jumps in stock returns increases from once every four months in the JDO model (Table 2) to once every two

Table 3

Parameter estimates for the stochastic volatility, jump–diffusion model

This table presents parameter estimates (standard errors) for the stochastic volatility, jump–diffusion model of stock prices given by

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{v_t} dW_{s,t} + [\exp(J_u) - 1] dN_u(\lambda_u) - [\exp(-J_d) - 1] dN_d(\lambda_d)$$

$$dv_t = \kappa(\theta - v_t) dt + \sigma\sqrt{v_t} dW_{v,t},$$

where the jump magnitudes J_u and J_d are draws from exponential distributions with means η_u and η_d , respectively, and $\text{Corr}(dW_{s,t}, dW_{v,t}) = \rho$.

The column “Daily” represents parameter estimates using daily stock price data for 1980–2000, the column “Weekly” represents estimates using weekly data for 1962–2000, while the column “Monthly” represents estimates using monthly data for 1926–1997.

Parameter	Daily data	Weekly data	Monthly data
μ	0.2238 (0.0540)	0.2153 (0.0447)	0.1581 (0.0281)
η_u	0.0513 (0.0093)	0.0481 (0.0178)	0.0346 (0.0207)
λ_u	0.3722 (0.0945)	0.1851 (0.0733)	0.0311 (0.0168)
η_d	0.0258 (0.0039)	0.0202 (0.0060)	0.0341 (0.0113)
λ_d	5.3791 (0.8900)	6.6105 (2.7273)	1.2399 (0.4910)
κ	12.1865 (3.9180)	3.8406 (1.7078)	0.6971 (0.4680)
θ	0.0185 (0.0021)	0.0203 (0.0021)	0.0314 (0.0046)
σ	4.2739 (0.7502)	0.5952 (0.1804)	0.3510 (0.0498)
ρ	−0.2712 (0.0849)	−0.4285 (0.1806)	−0.0994 (0.0700)

months in the SVJD model (Table 3) with both weekly and daily data. For monthly data this frequency drops, though not in a statistically significant sense, from the JDO model to the SVJD model, but the average magnitude of a negative jump increases from 1.8% to 3.4% per year. This suggests that negative jumps are an important component of stock returns even in the monthly data.

It is also important to note that ρ , the correlation coefficient between shocks to volatility and shocks to stock returns, drops (in absolute value) from −67% (Table 1) to −10% (Table 3) in monthly data after allowing for jumps in the process for stock returns. A similar drop is estimated with the daily data as well. The correlation parameter captures skewness in stock returns in the stochastic volatility only (SVO) model. This suggests that negative jumps, rather than stochastic volatility, drive the negative skewness that characterize stock returns, though stochastic volatility still remains important in explaining excess kurtosis (this can be seen from the fact that σ remains strongly significant in the SVJD model).

The results found here are consistent with those in other papers. For example, Andersen et al. (2002) find the specifications with both jumps and volatility to be the most robust in explaining daily returns both over the recent 20 years as well as the last 50 years. Their model loads on both the volatility correlation parameter as well as the jump parameter. In their paper, because the mean jump size is restricted to zero, the volatility correlation parameter picks up the negative skewness in returns, while jumps are found to be necessary in addition to time-varying volatility to explain excess kurtosis. This result is true in Pan (2002) as well. However, because the mean jump size is not restricted to be zero in her model, jumps do explain a portion of the negative skewness. However, similar to our results, the correlation parameter associated with volatility plays the bigger part in explaining negative skewness. A similar result is found in Chernov et al. (2002). Their estimate of the mean jump size is negative, so it helps explain some of the negative skewness in the data, but the mean is fairly small. The negative skewness in the data loads heavily on the correlation parameter associated with volatility. Similar results are found in Chernov et al. (2002). CGGT (2002) also estimate models that incorporate jumps in volatility as well as stock returns. This type of model has the appealing feature that when negative jumps hit returns, they can result in positive jumps in volatility. The high persistence of volatility then allows volatility to stay high for a short while—creating a kind of high volatility regime, which is a popular notion. However, upon estimating this model, there does not seem to be clear evidence of the superiority of this model specification versus one which incorporates jumps only in returns.

5.4. Non-affine models: non-affine stochastic volatility

We have so far shown applications of spectral GMM to affine models for stochastic processes—i.e., to stochastic processes whose characteristic function is log-linear in the state variables. But spectral GMM estimation works for any type of stochastic process—affine or non-affine, discrete-time or continuous-time. In this section, we show one application of the spectral GMM methodology to an interesting problem in finance that involves a non-affine stochastic process for stochastic volatility. This application involves constructing an approximate affine characteristic function for the underlying process using perturbation methods. In particular, we want to estimate the following generalized version of the stochastic volatility model (12)–(13):

$$\begin{aligned} \frac{dS_t}{S_t} &= \mu dt + \sqrt{v_t} dW_S, \\ dv_t &= \kappa(\theta - v_t) dt + \sigma v_t^{\gamma/2} dW_v, \end{aligned} \quad (23)$$

where W_S and W_v have an instantaneous correlation of ρ .

This model generalizes the stochastic volatility model (12)–(13) by allowing the instantaneous standard deviation of variance to be proportional to any power of variance.²² For values of γ other than one, this results in a non-affine stochastic volatility model because the square of the diffusion term for the volatility process

²² Model (12)–(13) restricts γ to be equal to 1.

is no longer linear in v_t . Chan et al. (1992) have proposed a model like (23) for the instantaneous interest rate, and they have found an estimate of γ equal to 3. The literature on volatility estimation has not yet estimated these types of models simply because estimation of even affine models have been so inherently difficult. However, this estimation problem is relatively simple using perturbation methods and spectral GMM.

In order for a stochastic differential equation to be well-behaved (for moments to exist, discretization schemes to converge weakly, and for strong solutions to exist), it needs to be square-integrable. To ensure this condition, it needs to satisfy local Lipschitz and linear growth conditions (see Karatzas and Shreve, 1988). A necessary condition for satisfying the growth condition is that γ needs to be less than 2. It is common, however, in the finance literature to estimate the value of γ without imposing this restriction (e.g., Chan et al. 1992; Aït-Sahalia, 1996). Since we are simply trying to demonstrate a methodology for estimating non-affine models, we will also follow the literature and not impose this restriction.²³ When interpreting the results, therefore, care should be taken in assigning too much weight on values of γ estimated to be greater than 2. An estimate of γ significantly greater than 2 should only be interpreted as a rejection of our choice of model specification, (23), rather than the possibility of the data generating process being of the form in (23) with $\gamma > 2$.

To apply spectral GMM, the first step is to calculate the conditional characteristic function. We do this for the transformed process

$$d \log S_t = (\mu - \frac{1}{2} v_t) dt + \sqrt{v_t} dW_S,$$

$$dv_t = \kappa(\theta - v_t) dt + \sigma v_t^{\gamma/2} dW_v.$$

From (4), the conditional characteristic function, $\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t)$, satisfies the following PDE:

$$\mathcal{D}\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t, v_t) = 0,$$

where

$$\begin{aligned} \mathcal{D}\phi = & \frac{1}{2} v_t \frac{\partial^2 \phi}{\partial \log S_t^2} + \rho \sigma v_t^{(\gamma+1)/2} \frac{\partial^2 \phi}{\partial \log S_t \partial v_t} + \frac{1}{2} \sigma^2 v_t^\gamma \frac{\partial^2 \phi}{\partial v_t^2} \\ & + \left(\mu - \frac{1}{2} v_t \right) \frac{\partial \phi}{\partial \log S_t} + \kappa(\theta - v_t) \frac{\partial \phi}{\partial v_t} - \frac{\partial \phi}{\partial \tau}. \end{aligned}$$

The boundary condition for (15) is given by $\phi(\omega, 0; \boldsymbol{\theta}, \log S_T, v_T) = \exp(i\omega \log S_T)$.

This is no longer a linear PDDE. The general results stated earlier for affine processes no longer apply here, and there is no known exact analytical solution to this equation.

²³ It is important to note that the mere fact that γ can be estimated to be greater than 2 indicates the level of inaccuracy that the affine approximation can lead to. Care needs to be taken in general when interpreting estimates from linearized non-affine models as parameters can be badly biased.

However, we utilize a perturbation method to derive an approximate solution.²⁴ The main feature of the method relies on approximating $v_t^{(\gamma+1)/2}$ and v_t^γ in the PDDE using Taylor expansions around the unconditional mean of v_t as follows:

$$v_t^{(\gamma+1)/2} \approx \theta^{(\gamma+1)/2} \left(\frac{1-\gamma}{2} \right) + \frac{\gamma+1}{2} \theta^{(\gamma-1)/2} v_t,$$

$$v_t^\gamma \approx \theta^\gamma (1-\gamma) + \gamma \theta^{\gamma-1} v_t.$$

These approximations result in the following PDDE:

$$\begin{aligned} 0 = & \frac{1}{2} v_t \frac{\partial^2 \phi}{\partial \log S_t^2} + \rho \sigma \left[\theta^{(\gamma+1)/2} \left(\frac{1-\gamma}{2} \right) + \frac{\gamma+1}{2} \theta^{(\gamma-1)/2} v_t \right] \frac{\partial^2 \phi}{\partial \log S_t \partial v_t} \\ & + \frac{1}{2} \sigma^2 [\theta^\gamma (1-\gamma) + \gamma \theta^{\gamma-1} v_t] \frac{\partial^2 \phi}{\partial v_t^2} + \left(\mu - \frac{1}{2} v_t \right) \frac{\partial \phi}{\partial \log S_t} \\ & + \kappa (\theta - v_t) \frac{\partial \phi}{\partial v_t} - \frac{\partial \phi}{\partial \tau}. \end{aligned}$$

This equation has an exponential-affine solution of the form,

$$\phi(\omega, \tau; \boldsymbol{\theta}, \log S_t, v_t) = \exp[i\omega \log S_t + A(\omega, \tau; \boldsymbol{\theta})v_t + B(\omega, \tau; \boldsymbol{\theta})], \tag{24}$$

where A and B solve the following two ordinary differential equations:

$$\frac{dA}{d\tau} = \frac{1}{2} \sigma^2 \gamma \theta^{\gamma-1} A^2 + \left[\rho \sigma i \omega \frac{\gamma+1}{2} \theta^{(\gamma-1)/2} - \kappa \right] A - \frac{1}{2} i \omega (i \omega - 1), \tag{25}$$

$$\frac{dB}{d\tau} = \frac{1}{2} \sigma^2 \theta^\gamma (1-\gamma) A^2 + \left[\rho \sigma i \omega \theta^{(\gamma+1)/2} \left(\frac{1-\gamma}{2} \right) + \kappa \theta \right] A + \mu i \omega. \tag{26}$$

Solving these linear ODEs is fairly simple to do. Subsequently, we need to integrate the instantaneous variance out of the solution to make the characteristic function conditional only on the observed stock price, just as we did in Section 5.1.

Alternatively, we can use the results in Section 5.1 directly by noting that the conditional characteristic function in (24) is equivalent to the conditional characteristic function for the stochastic volatility model (14) with $\theta, \sigma,$ and ρ replaced with $\tilde{\theta}, \tilde{\sigma},$ and $\tilde{\rho}$:

$$\tilde{\theta} = \theta \left[\rho \sigma \theta^{(\gamma-1)/2} \left(\frac{1-\gamma}{2} \right) + \frac{1}{2} \sigma^2 \theta^{\gamma-1} (1-\gamma) A + 1 \right],$$

$$\tilde{\sigma}^2 = \sigma^2 \gamma \theta^{\gamma-1},$$

$$\tilde{\rho} = \rho \frac{\gamma+1}{2\sqrt{\gamma}}.$$

²⁴ See Kevorkian and Cole (1981) for more on perturbation methods.

Table 4

Parameter estimates for the non-affine stochastic volatility model

This table presents parameter estimates (standard errors) for the non-affine stochastic volatility model of stock prices given by

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{v_t} dW_{s,t},$$

$$dv_t = \kappa(\theta - v_t) dt + \sigma v_t^{\gamma/2} dW_{v,t},$$

where $\text{Corr}(dW_{s,t}, dW_{v,t}) = \rho$.

The column “Daily” represents parameter estimates using daily stock price data for 1980–2000, the column “Weekly” represents estimates using weekly data for 1962–2000, while the column “Monthly” represents estimates using monthly data for 1926–1997.

Parameter	Daily data	Weekly data	Monthly data
μ	0.1196 (0.0393)	0.1096 (0.0210)	0.1248 (0.0226)
κ	13.7259 (3.0526)	4.5040 (3.1088)	0.7225 (0.4003)
θ	0.0341 (0.0026)	0.0255 (0.0034)	0.0347 (0.0042)
σ	4.5403 (0.4191)	0.7893 (0.1312)	1.5035 (0.2974)
ρ	−0.5089 (0.1845)	−0.3218 (0.1848)	−0.7864 (0.5327)
γ	1.8927 (0.5041)	2.1087 (0.7149)	2.9480 (0.9825)

Hence we can use the expressions in Section 5 to estimate (23) using spectral GMM on this approximate characteristic function.

Table 4 shows the results of this estimation using the same weekly and monthly data sets as in the previous applications. The point estimates and standard errors are adjusted for any bias induced by the approximation using a bootstrap procedure (explained below). The conclusion is similar to that obtained by Chan et al. (1992) in the context of interest rates. The value of γ appears to be different from 1.0, the baseline case that leads to an affine volatility model. With monthly data, our estimate of γ is 3.3, while with weekly data our estimate of γ is 2.2. The standard errors indicate that the differences between these values and the baseline value for γ of one are statistically significant. This suggests that the affine volatility model might not be a good description of stock return data; in fact, due to the regularity conditions restricting the value of γ , it suggests that the entire specification in (23) should perhaps be rejected.

However, this estimation exercise does not incorporate jumps into the stock price process. If, as suggested by the estimation results in Section 5.3, both stochastic volatility and jump diffusions are important in explaining strong negative skewness and excess kurtosis in stock returns, the large estimate of γ might be just the result of forcing the model to ignore the jump components. The inclusion of jumps may lessen the point estimate for γ , as jumps may account for the strong negative skewness and some of the excess kurtosis in the data.

Table 5

Parameter estimates for the non-affine stochastic volatility, jump–diffusion model

This table presents parameter estimates (standard errors) for the non-affine stochastic volatility model of stock prices given by

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{v_t} dW_{s,t} + [\exp(J_u) - 1] dN_u(\lambda_u) - [\exp(-J_d) - 1] dN_d(\lambda_d),$$

$$dv_t = \kappa(\theta - v_t) dt + \sigma v_t^{\gamma/2} dW_{v,t},$$

where $\text{Corr}(dW_{s,t}, dW_{v,t}) = \rho$.

The column “Daily” represents parameter estimates using daily stock price data for 1980–2000, the column “Weekly” represents estimates using weekly data for 1962–2000, while the column “Monthly” represents estimates using monthly data for 1926–1997.

Parameter	Daily data	Weekly data	Monthly data
μ	0.2205 (0.0377)	0.2296 (0.0438)	0.1618 (0.0279)
η_u	0.0316 (0.0063)	0.0309 (0.0078)	0.0284 (0.0147)
λ_u	0.4059 (0.0407)	0.2362 (0.0756)	0.0197 (0.0112)
η_d	0.0188 (0.0032)	0.0156 (0.0043)	0.0279 (0.0095)
λ_d	6.0713 (1.3698)	6.3912 (2.5793)	1.3726 (0.4327)
κ	11.8948 (3.0980)	4.9220 (1.9892)	0.5304 (0.3031)
θ	0.0223 (0.0015)	0.0213 (0.0016)	0.0327 (0.0039)
σ	4.7199 (0.6641)	0.5280 (0.1608)	0.3971 (0.0822)
ρ	-0.1926 (0.0714)	-0.3796 (0.1923)	-0.2329 (0.1198)
γ	1.2193 (0.2589)	1.1884 (0.3305)	1.7138 (0.6167)

In order to test the influence of misspecification on the estimate of γ obtained in Table 4, we re-estimated the model with jumps included in the model:

$$d \log S_t = \left(\mu - \frac{1}{2} v_t \right) dt + \sqrt{v_t} dW_S + [\exp(J_u) - 1] dN_u(\lambda_u) + [\exp(-J_d) - 1] dN_d(\lambda_d),$$

$$dv_t = \kappa(\theta - v_t) dt + \sigma v_t^{\gamma/2} dW_v, \tag{27}$$

where the jumps are modeled as in previous section. The jump densities are given by (19). The characteristic function for this process is a straightforward extension of the one used without jumps in this section.²⁵ Table 5 displays the estimation

²⁵ The characteristic function is extended in exactly the same way the affine stochastic volatility model was extended when jumps were added in the previous section. The same three additional terms appear in the exponential function.

results.²⁶ What is noteworthy about these results is the fact that the estimate for γ is now no longer statistically different from the value of 1.0. These results seem to indicate that the higher values of γ obtained previously were likely due to the omission of jumps from the model. While these results are only suggestive, they do seem to parallel those found in the interest rate literature, where the addition of jumps also seem to reduce the Chan et al. (1992) estimates for γ .

The purpose of this section was to illustrate one way that spectral GMM could be used with non-affine stochastic processes. With a simple perturbation, we were able to produce approximate characteristic functions for non-linear processes that could be used to generate moment restrictions in the spectral GMM procedure. However, it should be noted that the point estimates produced through any type of approximation method are biased and inconsistent. A theoretical value for the bias is difficult to calculate, so we used a bootstrap approach to partially correct for the bias.²⁷ As a rough check to see how far off our results might be, we simulated the model in (23) using a known set of parameters. 10,000 data sets of 30 years of weekly data were created. Table 6 displays these results. The effect of the approximation does not seem to be statistically significant for most parameters as the means of the parameter distributions are within roughly 1 standard deviation of the true estimates, including the parameter estimate for γ . The exception occurs with the estimate for θ . This parameter estimate, 0.0357, is significantly different from the true value of 0.04. The difference is likely a reflection of the bias caused by the approximation.

5.5. Volatility persistence

Our final application applies spectral GMM method to offer one possible explanation to a puzzling phenomenon observed in many financial markets. It has been observed that in estimating volatility, the point estimate for the rate of mean reversion changes dramatically with the frequency of the observed data. For example, Table 1 shows that for the stochastic volatility model in (12) above, the point estimate of the rate of mean reversion, κ , changes from 0.6 to 2.5 to 14 as we go from monthly to weekly to daily data. This feature has been observed in many financial markets for volatility and non-volatility processes alike. The point estimates for the rate of mean reversion drop

²⁶ As with the estimates in Table 4, these estimates have been modified by a bootstrap method to adjust for the bias caused by the approximation.

²⁷ The procedure we used was to simply simulate sets of data using the initial values for the estimated parameters. Then, parameter estimates are produced for each of these new data sets. This results in a distribution for each parameter. The difference between the mean of this distribution and the initial parameter estimates gives a rough value for the bias in the initial parameter estimates. The initial parameter estimate is then adjusted using this estimate for the bias. While this procedure is crude, we have found this bootstrap procedure to reduce the bias inherent in parameter estimates produced from approximation-based methods (such as the perturbation method used in this paper as well as the commonly used Euler approximation) considerably. See Efron and Tibshirani (1993) for more on the bootstrap.

Table 6

Measure of error produced via perturbation method

This table presents the results of simulating the stochastic volatility model

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{v_t} dW_{s,t},$$

$$dv_t = \kappa(\theta - v_t) dt + \sigma v_t^{\gamma/2} dW_{v,t}$$

to produce 10,000 data sets of 30 years of weekly stock return data. The parameters used in the simulations are given below under the heading “simulation value”. Then, these datasets are used to perform an approximate estimation using a perturbation approach. The mean and standard deviation of the estimated parameter distributions are given in far right column.

Parameter	Simulation value	Estimated value
μ	0.1000	0.1103 (0.3696)
κ	1.000	0.6896 (0.4265)
θ	0.0400	0.0357 (0.0021)
σ	1.500	1.3340 (0.1218)
ρ	-0.5000	-0.4258 (0.2672)
γ	1.500	1.8128 (0.6211)

as the frequency of the data drops, i.e., persistence in volatility appears to increase when lower frequency data is used for estimation.

In this section, we offer one possible explanation. We speculate that there might be multiple frequency components to volatility, and that lowering the frequency of the data used for estimation simply causes a single-factor stochastic volatility model to load on the lower frequency (higher persistence) volatility component.

We begin by assuming that the data generating process for a stock price is given by

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{v_t} dW_S,$$

where the instantaneous variance is determined by three additive components:

$$v_t = x_t + y_t + z_t.$$

Each component is determined by a square-root process:

$$dx_t = \kappa_x(\theta - x_t) dt + \sigma \sqrt{x_t} dW_v,$$

$$dy_t = \kappa_y(\theta - y_t) dt + \sigma \sqrt{y_t} dW_v,$$

$$dz_t = \kappa_z(\theta - z_t) dt + \sigma \sqrt{z_t} dW_v.$$

Table 7

Parameter estimates for multiple factor stochastic volatility model

This table presents parameter estimates for the stochastic volatility model of stock prices given by

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{v_t} dW_{s,t},$$

$$dv_t = \kappa(\theta - v_t) dt + \sigma\sqrt{v_t} dW_{v,t},$$

where $\text{Corr}(dW_{s,t}, dW_{v,t}) = \rho$. However, the data generating process is a multiple factor volatility model given by

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{v_t} dW_{s,t}.$$

Volatility is composed of three factors: $v_t = x_t + y_t + z_t$. The three factors are themselves determined by square root processes with different rates of mean reversion.

$$dx_t = \kappa_x(\theta - x_t) dt + \sigma\sqrt{x_t} dW$$

$$dy_t = \kappa_y(\theta - y_t) dt + \sigma\sqrt{y_t} dW$$

$$dz_t = \kappa_z(\theta - z_t) dt + \sigma\sqrt{z_t} dW.$$

The parameters used for the data generating process were $\mu = 0.13$, $\kappa_x = 0.2$, $\kappa_y = 1$, $\kappa_z = 5$, $\theta = 0.01$, and $\sigma = 0.05$.

The columns “Weekly”, “Monthly”, and “Annual” represent parameter estimates using weekly, monthly, and annual sampling frequencies for the data produced from the data generating process. 100 data sets were generated, and the parameter estimates below give the means of the point estimates produced with each dataset.

Parameter	Weekly data	Monthly data	Annual data
μ	0.1715	0.1191	0.1380
κ	5.3315	2.7060	0.7601
θ	0.0282	0.0276	0.0348
σ	0.3356	0.2550	0.2188
ρ	-0.8738	-0.5815	-0.3550

The instantaneous correlation between dW_S and dW_v is given by ρ . Note that the only difference between each component is in the rate of mean reversion, and they are all subject to the same shock.²⁸ Thus, each process represents a different frequency component of volatility.

For our specific example, we parameterize the data generating process as follows: $\mu = 0.13$, $\kappa_x = 0.2$, $\kappa_y = 1$, $\kappa_z = 5$, $\theta = 0.01$, $\sigma = 0.05$ and $\rho = -0.5$. To show how sampling frequency affects estimation, we generate data from this process and estimate the parameters of the process via spectral GMM by sampling this data at different frequencies and using the single factor model (12)–(13). We generate 100 data sets to ensure that one unusual sample path for the stock price does not skew the results.

Table 7 reports the means of the estimates from the 100 sets of parameter estimates. As we speculated, the estimate of κ , the rate of mean reversion, decreases as we decrease the sampling frequency from weekly to monthly to annual. The estimate of κ

²⁸ Note that the results presented here also work when θ and σ differ across the volatility components. We could also allow for a different shock to each component. However, we keep things as simple as possible to illustrate our point.

starts from 5.33 with weekly data, decreases to 2.71 with monthly data, and it further decreases to 0.76 with annual data. This is precisely the pattern observed for the rate of mean reversion in volatility in many financial markets—Table 1 illustrates this pattern for the US equity market. As we lower the sampling frequency from weekly to monthly to annual, the single factor model is forced to load on one of the volatility factors of the data generating process. The model loads on the volatility factor whose frequency is closest to the sampling frequency. Therefore, we observe the rate of mean reversion decreasing, and persistence increasing, as we decrease the sampling frequency. Other interesting results to note is that the correlation level, ρ , also drops (in absolute value) as the sampling frequency drops, while the parameter σ displays a strong positive bias due to the model misspecification.

Chernov et al. (2000, 2002) estimate multi-factor stochastic volatility models as well, though they do not explore this changing volatility persistence effect explored in this section. Chernov et al. (2000, 2002) find that the addition of a second volatility factor improves the fit of the model relative to a single-factor volatility model. However, when the choice of using jumps or a second volatility factor in the model needs to be made, it appears that jumps are far more important in capturing return dynamics than the second volatility factor. This could be indicative that the effects discussed in this section are second-order in comparison to the effects of incorporating jump processes in returns. Additionally, Chernov et al. (2002) find that when a second volatility factor is allowed to have its own correlation with returns (we impose a constant correlation between all volatility factors and the returns process), the correlation parameters can take on both positive and negative values, contrary to the findings in single factor volatility models, where the correlation parameter is always found to be negative.

While the results of this section are far from conclusive, they are suggestive of one potential explanation for the pattern of persistence observed in volatility and also interest rates and exchange rates. There could be different frequency components to each of these financial variables, but because researchers tend to model these as one factor models, they pick up only a narrow band of these frequencies depending on the particular sampling frequency used for the data. Indeed, in the case of interest rates, researchers have determined using principal components and other statistical analyses that there are three important factors determining interest rates. We suspect that similar results will hold true for volatility in many financial markets as well.

6. Conclusion

This paper derives a methodology for the direct estimation of continuous-time stochastic models based on the characteristic function. The estimation method does not require discretization of the process, and it is easy to apply. The method is essentially generalized method of moments on the complex plane. Hence it shares the optimality and distribution properties of GMM estimators. Moreover, an appropriate choice of instruments delivers an asymptotically efficient estimator. This estimation method expands the set of continuous-time stochastic processes for which simple estimation without

discretization is feasible. This is so because computing the characteristic function is easier than computing the likelihood function for a large number of stochastic processes.

We illustrate the method with some applications to relevant estimation problems in continuous-time finance. We estimate a model of stochastic volatility, and we show that stochastic volatility is important in capturing stock return dynamics. We also estimate a jump–diffusion model with constant volatility and show that both upward and downward jumps are also important in explaining stock returns. Indeed, both the stochastic volatility model and the jump–diffusion model can capture the skewness and excess kurtosis that we observe in stock returns. Next we estimate a model that nests both the stochastic volatility model and the jump–diffusion model to ascertain the contribution of each component in explaining the higher order moments in stock returns. We find that negative jumps are important to explain skewness and asymmetry in excess kurtosis of the return distribution, while stochastic volatility is important to capture the overall level of this kurtosis. Positive jumps are not statistically significant once we allow for stochastic volatility in the model. We also explore a potential explanation for the observation that the point estimate for the rate of mean reversion in a stochastic volatility model decreases dramatically with the frequency of the observed data. We show that this is consistent with a model of multiple additive components in volatility, each of them operating at a different frequency.

Most of the processes in the paper have characteristic functions which are exponential affine in the vector of state variables (mirroring the current practice in finance), but this method is also feasible in non-affine settings if one utilizes perturbation methods. To illustrate this, we estimate a non-affine model of stochastic volatility with an arbitrary power in the diffusion coefficient. We find that the power of the diffusion coefficient appears to be between one and two, rather than the value of one-half that leads to the standard affine stochastic volatility model. However, we also show that this result may be driven by model misspecification. When we include jumps into this non-affine model, the power of the diffusion coefficient is no longer estimated to be statistically different from one-half.

The estimates we present in this paper are based on a finite set of moment conditions implied by the empirical characteristic function of the process at hand. However, the characteristic function of a process generates a continuum of moment conditions. In a recent paper, Carrasco et al. (2001), building on the work of Carrasco and Florens (2000), utilize a continuum of moment conditions to estimate multivariate diffusions, and show that this improves the efficiency of the estimator, as this allows the estimator to achieve the lower Cramer–Rao bound. Incorporating the full set of infinite moment conditions into the spectral GMM procedure would improve the efficiency of the estimator.

Acknowledgements

We want to thank participants in seminars at Boston University, Cornell University, Harvard University, the 2000 NBER Summer Institute, and the Western Finance Association Meetings for their comments and suggestions. We have also benefitted from

discussions with Torben Andersen, John Campbell, Mikhael Chernoff, Sanjiv Das, Ken Froot, Christopher Jones, and Ken Singleton. Finally, the editors of this issue, and two anonymous referees have given us extremely valuable suggestions which have greatly improved this paper. We also thank the Division of Research at the Harvard Business School for financial support.

References

- Ahn, C., Thompson, H., 1988. Jump–diffusion processes and the term structure of interest rates. *Journal of Finance* 43, 155–174.
- Ait-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. *Review of Financial Studies* 9, 385–426.
- Amin, K., Ng, V., 1993. Option valuation with systematic stochastic volatility. *Journal of Finance* 48, 881–910.
- Andersen, T.L., Benzoni, J., Lund, 2002. An empirical investigation of continuous-time equity return models. *Journal of Finance* 57, 1239–1284.
- Bakshi, G., Madan, D., 2000. Spanning and derivative security valuation. *Journal of Financial Economics* 55, 205–238.
- Bates, D., 1996. Jumps and stochastic volatility: exchange rate processes implicit in deutsche mark options. *Review of Financial Studies* 9, 69–107.
- Black, F., 1976. Studies of stock price volatility changes. in: *Proceedings of the 1976 Meetings of the Business and Economic Statistics Section, American Statistical Association*, pp. 177–181.
- Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81, 637–654.
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Bollerslev, T., Chou, R., Kroner, K., 1992. ARCH modelling in finance. *Journal of Econometrics*, 52, 5–59.
- Carrasco, M., Florens, J.P., 2000. Generalization of GMM to a continuum of moment conditions. *Econometric Theory* 16, 797–834.
- Carrasco, M., Chernov, M., Florens, J.P., Ghysels, E., 2001. Estimating diffusions with a continuum of moment conditions. Working paper.
- Chacko, G., 1999. Estimating exponential separable term structure models: a general approach. Unpublished Manuscript, Harvard University.
- Chacko, G., Das, S., 2002. Pricing interest rate derivatives: a general approach. *Review of Financial Studies* 15, 195–241.
- Chacko, G., Viceira, L., 1999. Dynamic consumption and portfolio choice with stochastic volatility in incomplete markets. NBER Working paper 7377.
- Chan, K., Karolyi, A., Longstaff, F., Sanders, A., 1992. An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance* 47, 1209–1227.
- Chernov, M., Ghysels, E., 2000. A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *Journal of Financial Economics* 56, 407–458.
- Chernov, M., Gallant, R., Ghysels, E., Tauchen, G., 2000. A new class of stochastic volatility models with jumps: theory and estimation. Working paper.
- Chernov, M., Gallant, R., Ghysels, E., Tauchen, G., 2002. Alternative models for stock price dynamics. Working paper.
- Danielsson, J., 1994. Stochastic volatility in asset prices: estimation with simulated maximum likelihood. *Journal of Econometrics* 54.
- Das, S., 1996. Poisson–Gaussian processes and the bond markets. Working paper, Harvard University.
- Das, S., Foresi, S., 1996. Exact solutions for bond and options prices with systematic jump risk. *Review of Derivatives Research* 1, 7–24.
- Das, S., Sundaram, R., 1999. Of smiles and smirks: a term structure perspective. *Journal of Financial and Quantitative Analysis* 34, 60–72.

- Duffie, D., Kan, R., 1993. A yield factor model of interest rates. *Mathematical Finance* 6, 379–406.
- Duffie, D., Pan, J., Singleton, K., 2000. Transform analysis and option pricing for affine jump–diffusions. *Econometrica* 68, 1343–1376.
- Efron, B., Tibshirani, R., 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- Engle, R., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50, 987–1008.
- Feuerverger, A., 1990. An efficiency result for the empirical characteristic function in stationary time-series models. *The Canadian Journal of Statistics* 18, 155–161.
- Feuerverger, A., 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.
- Gallant, R., Tauchen, G., 1996. Which moments to match? *Econometric Theory* 12, 657–681.
- Gallant, A.R., Hsieh, D., Tauchen, G., 1997. Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics* 81, 159–192.
- Ghysels, E., Harvey, A., Renault, E., 1996. Stochastic volatility. In: Madala, G.S., Rao, C.R. (Eds.), *Handbook of Statistics*, Vol. 14. Elsevier, Amsterdam.
- Grimmet, G.R., Stirzaker, D.R., 1992. *Probability and Random Processes*. Oxford University Press, New York.
- Hansen, L., 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.
- Harvey, A., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. *Review of Economic Studies* 61, 247–264.
- Hentschel, L., 1995. All in the family: nesting symmetric and asymmetric GARCH models. *Journal of Financial Economics* 39, 71–104.
- Heston, S., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6, 327–343.
- Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatilities. *Journal of Finance* 42, 281–300.
- Jacquier, E., Polson, N., Rossi, P., 1994. Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* 12, 371–389.
- Jiang, G., Knight, J., 2002. Estimation of continuous time processes via empirical characteristic function. *Journal of Business and Economic Statistics* 20, 198–212.
- Jorion, P., 1988. On jump processes in the foreign exchange and stock markets. *Review of Financial Studies* 1, 427–445.
- Karatzas, I., Shreve, S., 1988. *Brownian Motion and Stochastic Calculus*. Springer, New York.
- Kevorkian, J., Cole, J.D., 1981. *Perturbation Methods in Applied Mathematics*. Springer, New York.
- Lo, A., 1988. Maximum likelihood estimation of generalized Ito processes with discretely sampled data. *Econometric Theory* 4, 231–247.
- Melino, A., Turnbull, S., 1990. Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* 45, 239–265.
- Merton, R., 1973. The theory of rational option pricing. *Bell Journal of Economics and Management Science* 4, 141–183.
- Merton, R., 1976. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* 3, 125–144.
- Nelson, D., 1989. Modelling stock market volatility changes. *ASA 1989 Proceedings of the Business and Economic Statistics*, 93–98.
- Nelson, D., 1990. ARCH Models as diffusion approximations. *Journal of Econometrics* 45, 7–38.
- Newey, W., West, K., 1987. A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Pan, J., 2002. The jump-risk premia implicit in options: evidence from an integrated time-series study. *Journal of Financial Economics* 63, 3–50.
- Silvey, S.D., 1975. *Statistical Inference*. Chapman and Hall, New York.

- Singleton, K., 2001. Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics* 102, 111–141.
- Stein, E., Stein, J., 1991. Stock price distributions with stochastic volatility: an analytical approach. *Review of Financial Studies* 4, 727–753.
- Wiggins, J., 1987. Option values under stochastic volatility: theory and empirical estimates. *Journal of Financial Economics* 19, 351–372.