Annals of Finance
DOI 10.1007/s10436-006-0062-y

RESEARCH ARTICLE

Maximum likelihood estimation of the double exponential jump-diffusion process

Cyrus A. Ramezani · Yong Zeng

Received: 21 July 2005 / Accepted: 2 October 2006

Abstract The double exponential jump-diffusion (DEJD) model, recently proposed by Kou (Manage Sci 48(8), 1086–1101, 2002) and Ramezani and Zeng (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=606361, 1998), generates a highly skewed and leptokurtic distribution and is capable of matching key features of stock and index returns. Moreover, DEJD leads to tractable pricing formulas for exotic and path dependent options (Kou and Wang Manage Sci 50(9), 1178–1192, 2004). Accordingly, the double exponential representation has gained wide acceptance. However, estimation and empirical assessment of this model has received little attention to date. The primary objective of this paper is to fill this gap. We use daily returns for the S&P-500 and the NASDAQ indexes and individual stocks, in conjunction with maximum likelihood estimation (MLE) to fit the DEJD model. We utilize the BIC criterion to assess the performance of DEJD relative to log-normally distributed jump-diffusion (LJD) and the geometric brownian motion (GBM). We find that DEJD performs better than these alternatives for both indexes and individual stocks.

Keywords Asset price processes · Double exponential jump-diffusion · Pareto-beta jump diffusion · Leptokurtic distributions · Volatility smile-smirk · MLE

C. A. Ramezani (✉)
Finance area, Orfalea college of business, California Polytechnic State University, San Luis Obispo, CA 93407, USA
e-mail: cramezan@calpoly.edu

Y. Zeng
Department of Mathematics and Statistics, University of Missouri, Kansas City, MO 64110, USA
e-mail: zeng@mendota.umkc.edu

JEL Classification Numbers C32 · C52 · G12 · G13

1 Introduction

Three decades have passed since the seminal papers by Merton (1976a,b), suggested that asset price dynamics may be modeled as jump-diffusion (JD) processes and provided the foundations to value contingent claims under this specification. Building on the works of Press (1967), Merton posited that an asset's returns process may be decomposed into three components; a linear drift, a Brownian motion representing "normal" price variations, and a compound Poisson process that generates "news" arrivals leading to "abnormal" change in prices (jumps). Upon the arrival of news, jump magnitudes are determined by sampling from an independent and identically distributed (IID) random variable. For the purpose of pricing options, Merton assumed that the jump magnitude are log-normally distributed (LJD hereafter). This special case renders estimation and hypothesis testing tractable and has become the most important representation of the JD process.

In the extant literature, generalizations of the JD class of representations occur by assuming different theoretical structures for the drift, the diffusion, and the jump component of the process (e.g., stochastic volatility and mean reversion). Numerous variations have been proposed to enhance the jump specification, including different distributional assumptions for the jump magnitude, time varying jump intensity, and correlated jump magnitudes, to name a few. For example, Naik (1993) proposed a model in which return volatility jumps from one regime to another. Andersen and Andreasen (2000) combined a deterministic volatility structure with log-normally distributed Poisson jumps. In the end, a large number of continuous time models emerge, primarily by choosing different specifications for the basic building blocks. Each specific formulations can be shown to be a special case of the affine jump diffusion (AJD) framework of Duffie et al. (2000). Moreover, empirical evidence indicates that option pricing formulas based on variants of the AJD exhibit less bias – i.e. better explain the "smile" and "skew" across both moneyness and maturity; see the survey by Garcia et al. (2004).

The popularity of AJD framework is due to its modeling flexibility that captures important features of financial processes, and its technical tractability in deriving standard and extended transforms for option and bond pricing, as well as econometric estimation. Chernov et al. (2003), Eraker et al. (2003), and the papers in Aït-Sahalia and Hansen (2004) provide a complete survey of the most important developments in the AJD literature, focusing on the econometric issues. Huang and Wu (2004) and Carr and Wu (2004) provide a systematic catalog of the AJD alternatives and their properties. Bates (2003a) provides an assessment of the progress in this area and highlights remaining challenges.

The Double exponential jump diffusion (DEJD) is a special case of the AJD family but it has desirable properties for both pricing exotic options and econometric estimation. In its most popular form, the JD model has a single jump component that captures the impact of news on security prices. News that cause upward jump in prices – "good news" – and news that cause downward jump in prices – "bad news" – are not distinguished by their intensity or distributional characteristics. This potential limitation of the simple JD framework has led to two alternative specifications. Under Kou's (2002) DEJD specification, a single Poisson process with fixed intensity generates the jumps in prices, but the jump magnitudes are drawn from two independent exponential distributions. Ramezani and Zeng (1998) independently propose the Pareto-Beta jump-diffusion (PBJD), assuming that good and bad news are generated by two independent Poisson processes and jump magnitudes are drawn from the Pareto and Beta distributions. Below we show that the two models are closely related in that the parameters of one model can be exactly recovered from the other.[1]

The DEJD model has gained popularity primarily because its distribution is asymmetric and Leptokurtic. Furthermore, as Kou (2002), Sepp (2004) and Kou and Wang (2004) have shown, the DEJD model leads to nearly analytical option pricing formula for certain exotic and path dependent options. This is a significant advantage as most of the existing methods for pricing options under the JD processes are confined to plain vanilla European options. Because of these features, the DEJD has been used to model the price process for stocks indexes, as well as pricing credit risk instruments tied to firms' asset returns.

The literature based on DEJD is large and expanding. For example, Huang and Huang (2003) study the link between Corporate-Treasury yield spread and default risk by calibrating the DEJD specifications to the firms' *asset returns*. Carr and Wu (2004) and Huang and Wu (2004) show that the symmetrical form of the DEJD (see below) is a special case of the general time-changed Lévy process. They use the characteristic function methodology to price contingent claims via an efficient fast fourier transform (FFT) technique. This methodology enables the authors to select and test alternative option pricing models. Carr and Madan (1999), Lewis (2001), Broadie and Yamamoto (2003), and Cont and Tankov (2004) also use variations of the FFT method to develop an option pricing formula under DEJD specification. dH́alluin et al. (2004) develop robust numerical methods for pricing contingent claims under jump diffusion processes, with particular emphasis on DEJD. Using Laplace Transform, Sepp (2004) develops pricing formulas for *double barrier* options under DEJD. Lee (2006) undertakes a systematic study of the determinant of the shape of the implied volatility surface for a host of processes, including the DEJD. Cont and Tankov (2004) propose a non-parametric method for calibrating JD option pricing models, including the DEJD. Finally, Keppo et al. (2003) extend the DEJD by introducing stochastic volatility.

Despite this growing interest in DEJD as a model of security price processes, estimation and empirical assessment of this model has received little attention to date. In practice, most studies have arbitrarily selected "reasonable"

[1] Ramezani and Zeng (1998) propose their model from an econometric viewpoint. These papers are clearly complementary: While Ramezani and Zeng (1998) focus on the problem of parameter estimation, Kou (2002) and Kou and Wang (2004) develop the DEJD option pricing formula, which would require the estimated parameters as inputs.

parameter values, because as Huang and Huang (2003) observe "no study has estimated the parameters for this model". The primary objective of this paper is to remedy this deficiency by providing an empirical assessment of the DEJD specification. Using maximum likelihood estimation procedure, daily data for individual stocks and the S&P-500 and the NASDAQ composites, we undertake a comprehensive empirical evaluation of DEJD. We utilize the BIC criterion to assess the performance of DEJD relative to LJD and GBM. We find that DEJD performs better than these alternatives for both indexes and individual stocks.

2 The model

In this section we first present the PBJD model of Ramezani and Zeng (1998). The PBJD assumes that good and bad news are generated by two independent Poisson processes and jump magnitudes are drawn from the Pareto and the Beta distributions. We then identify the conditions under which PBJD reduces to DEJD specification, which has a single Poisson process generating news and two independent exponential distributions generating the up and down jump magnitudes.

There are several economic justifications for making a distinction between good and bad news. At a microeconomic level, Milgrom (1981) has formalized the notion of good and bad news and shown that such distinction plays an important role in rational expectation models that are the foundation of information economics. At the firm level, discontinuous up and down price jumps may be a consequence of significant changes in the operating and financial structure of the firm. Moreover, Kou (2002) argues that investors' sentiment in the form of under- and over-reaction, as documented in the behavioral finance literature, leads to differential response to good and bad news. Finally, Maheu and McCurdy (2004) show that expansionary and contractionary economic periods are accompanied with unequal frequency of good and bad news arrivals.

The separation of good from bad news implies that the range of values for the random percentage change in price must be constrained. Because stocks represent limited liability, the percentage change in price due to bad news must be bounded from below by -100%. Similarly, the percentage change in price due to arrival of good news must be positive. Because of these constraints, care must be taken when choosing a distribution for either up or down jump magnitudes. Under the PBJD, the jump magnitudes for good and bad news are drawn from the Pareto and Beta distributions, respectively. In addition to having the appropriate supports, these distributions lead to a tractable likelihood function and facilitate MLE.

Let $S(t)$ denote the price of stock at time t and assume that the price process can be represented by the following:

$$\frac{\mathrm{d}S(t)}{S(t-)} = \mu \mathrm{d}t + \sigma \mathrm{d}Z(t) + \sum_{j=u,d} \left(V^j_{N^j(\lambda^j t)} - 1 \right) \mathrm{d}N^j(\lambda^j t) \tag{1}$$

where μ and σ are the drift and volatility terms, $Z(t)$ is a standard Wiener process, V^j is the jump magnitude, and $N^j(\lambda^j)$ are independent Poisson processes with intensity parameters λ^j ($j = u, d$ represent up- and down-jumps, respectively). It is assumed that the up-jump magnitudes (V^u) are distributed Pareto(η_u) with density function

$$f_{V^u}(x) = \left(\frac{\eta_u}{x^{\eta_u+1}}\right) \quad V^u \geq 1,$$
$$\mathrm{E}(V^u) = \frac{\eta_u}{\eta_u - 1} \quad \text{and} \quad \sigma^2_{V^u} = \frac{\eta_u}{(\eta_u - 2)(\eta_u - 1)^2}.$$

Similarly, the down-jump magnitudes (V^d) are distributed Beta(η_d,1) with density function

$$f_{V^d}(x) = \eta_d x^{\eta_d - 1} \quad 0 < V^d < 1,$$
$$\mathrm{E}(V^d) = \frac{\eta_d}{(\eta_d + 1)} \quad \text{and} \quad \sigma^2_{V^d} = \frac{\eta_d}{(\eta_d + 2)(\eta_d + 1)^2}.$$

All jumps are assumed to be independent, which implies a mixture of Pareto-Beta distributions for jump magnitudes. The specification in (1) is a Lévy process. It can be decomposed as the sum of three independent components, a linear drift, a Brownian motion and a pure jump process. Moreover, it has stationary and independent increments and is continuous in probability.[2]

At this point it may be tempting to extend this specification to include stochastic volatility. However, this will be a departure from our stated objective to provide parameter estimates for DEJD. Keppo et al. (2003) have extend the DEJD to include stochastic volatility but these authors do not obtain MLE estimates for this specification. In general, it is difficult to obtain a closed form likelihood function for a model that nests jump structures like Eq. (1) and continuous time conditional Heteroskedasticity. Moreover, as Huang and Wu (2004) observe, the variation in return volatility can be generated either by variations in diffusion variance, or variations in the arrival rates of jumps, or a combination of the two. These authors argue that the evidence favors variations in jump intensity, which is partially achieved by the double jump formulation above.

The Doléans-Dade formula (Protter, 1991) provides an explicit solution for (1):

$$S(t) = S(0) \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma Z(t)\right\} \prod_{j=u,d} V^j(N(\lambda^j t)) \tag{2}$$

$$\prod_{j=u,d} V^j(N(\lambda^j t)) = \begin{cases} 1 & \text{if } N(\lambda^j t) = 0 \\ \prod_{i=1}^{N(\lambda^j t)} V_i^j & \text{if } N(\lambda^j t) = 1, 2, 3, \ldots \end{cases}$$

[2] For recent application of Lévy process in finance see Cont and Tankov (2003), Carr and Wu (2004) and Huang and Wu (2004) and their references.

Let $Y = \ln(V)$, then the s period rate of return, $r(s)$, can be written as:

$$r(s) = \left(\mu - \frac{1}{2}\sigma^2\right)s + \sigma Z(s) + \sum_{i=1}^{N_s^u} Y_i^u + \sum_{i=1}^{N_s^d} Y_i^d \tag{3}$$

where $N_s^j, j = u$ or d, denotes the number of good (bad) news over the time period s. In Appendix 1 we derive the density functions for $r(s)$ and show that it is a probability weighted mixture of normal and exponential distributions. The density function provides the basis for maximum likelihood estimation (MLE) described below.

The connection between the PBJD and DEJD can now be established as follows. Let $\lambda = \lambda_u + \lambda_d, p = \frac{\lambda_u}{\lambda}$, and $q = 1 - p$, then the distribution of jump magnitudes, $\tilde{V}$, is a probability weighted mixture of Pareto and Beta:

$$f_{\tilde{V}}(x) = p\frac{\eta_u}{x^{\eta_u+1}}\mathbf{I}_{\{x>1\}} + q\eta_d x^{\eta_d-1}\mathbf{I}_{\{0<x<1\}}; \quad \eta_u > 1, \eta_d > 0$$

where, setting $Y = \ln(\tilde{V})$ and noting that the distribution of the logarithm of Pareto and Beta is exponential (see Appendix 1), we obtain:

$$f_Y(y) = p\eta_u \mathrm{e}^{-\eta_u y}\mathbf{I}_{\{y\geq 0\}} + q\eta_d \mathrm{e}^{\eta_d y}\mathbf{I}_{\{y<0\}} \tag{4}$$

This is the distribution of the logarithm of the jump magnitudes under the DEJD, which has a jump intensity λ and Y has an IID mixture distribution of exponential(η_u) and exponential(η_d) with probabilities p and q, respectively. Note that from an inference perspective, both models have the same number of parameters to estimate: $\theta_{\text{DEJD}} = (\mu, \sigma, \lambda, p, \eta_u, \eta_d)$ and $\theta_{\text{PBJD}} = (\mu, \sigma, \lambda_u, \lambda_d, \eta_u, \eta_d)$. Having established the connection between these models, in the reminder of the paper we use "DEJD" to refer to both.

Merton's Log-normal Jump-Diffusion (LJD) model also has a single jump component with magnitude V distributed IID log-normal (α, β^2) and Poisson (λ) arrival rate. However, it is important to note that the LJD and DEJD specifications *are not* nested. Without the jump components, these models reduce to the standard GBM.

Table 1 presents the first four moments of returns for GBM, LJD, and the DEJD. Clearly, relative to GBM, both LJD and DEJD are capable of generating a higher peak, positive or negative skewness, and positive kurtosis and are therefore likely to better match the empirical returns distribution. However, a priori it is not clear whether DEJD performs better than LJD in this regard. An empirical examination is needed to address this issue.

Focusing on the jump specification in (4), three special cases can be delineated:

Case (1) Suppose $\eta_u = \eta_d = \eta$ and $\lambda_u = \lambda_d = \lambda$ (i.e., $p = 0.5$), then the distribution of jumps will be symmetrical with higher peak and positive

Table 1 Moments of returns

	GBM	JD	DEJD
$Er(s)$	$(\mu - \frac{1}{2}\sigma^2)s$	$(\mu - \frac{1}{2}\sigma^2 + \lambda\alpha)s$	$(\mu - \frac{1}{2}\sigma^2 + \frac{\lambda_u}{\eta_u} - \frac{\lambda_d}{\eta_d})s$
$\mathrm{Var}[r(s)]$	$\sigma^2 s$	$(\sigma^2 + \lambda(\beta^2 + \alpha^2))s$	$(\sigma^2 + 2\frac{\lambda_u}{\eta_u^2} + 2\frac{\lambda_d}{\eta_d^2})s$
Skewness	0	$\frac{\lambda\alpha^3}{(\sigma^2+\lambda(\beta^2+\alpha^2))^{1.5}s^{0.5}}$	$\frac{6(\frac{\lambda_u}{\eta_u^3}-\frac{\lambda_d}{\eta_d^3})}{(\sigma^2+2\frac{\lambda_u}{\eta_u^2}+2\frac{\lambda_d}{r_d^2})^{1.5}\sqrt{s}}$
Kurtosis	0	$\frac{\lambda(\alpha^4+3\beta^4)}{(\sigma^2+\lambda(\beta^2+\alpha^2))^2 s}$	$\frac{24(\frac{\lambda_u}{\eta_u^4}+\frac{\lambda_d}{\eta_d^4})}{(\sigma^2+2\frac{\lambda_u}{\eta_u^2}+2\frac{\lambda_d}{\eta_d^2})^2 s}$

The first four moments of s period returns $(r(s))$ under GBM and two Jump-Diffusion specifications. The second and third moments are defined as *Skewness* $E(X - EX)^3/[Var(X)]^{3/2}$ and *Kurtosis* $E(X - EX)^4/[Var(X)]^2 - 3$

kurtosis relative to normal. Tsay (2002, page 245) discusses the properties of this special case and notes that for finite samples, it will be difficult to distinguish this distribution from the student-t. Unlike the latter, however, the former is tractable analytically and can generate a higher probability concentration around its mean. As Tsay (2002), Huang and Huang (2003), and Carr and Wu (2004) show, this form of symmetry leads to simpler option pricing formulas.[3] Absent actual parameter estimates, this assumption appears innocuous and has been commonly invoked.

Case (2) Suppose $\eta_u = \eta_d = \eta$ and $\lambda_u \neq \lambda_d$, then relative to GBM, the distribution of $r(s)$ will be skewed and have excess kurtosis and the relative size of λ_u and λ_d will lead to negative or positive skewness.

Case (3) Suppose $\eta_u \neq \eta_d$ and $\lambda_u = \lambda_d = \lambda$, again the resulting $r(s)$ will be skewed and show excess kurtosis. However, the relative size of η_u and η_d will determine whether the distribution is negatively or positively skewed.

Note that relative to LJD, the DEJD provides additional econometric flexibility in the following sense: Under LJD, a single compound Poisson process derives the news arrival and both up and down jumps have the same arrival intensity and jump distribution. Under the DEJD, however, the above special cases may be distinguished.

Both the LJD and the DEJD can explain the widely documented volatility smile in the option pricing literature. In particular, Kou (2002), Kou and Wang (2004), and others have demonstrated that the excess kurtosis and skewness of the DEJD under the physical measure generates similar features in the risk

[3] A Fortran program for pricing European options for this special case of the DEJD is available from Professor Tsay's web page: `http://www.gsb.uchicago.edu/fac/ruey.tsay/teaching/fts/kou.f`. Mathematica code for pricing European options under the general DEJD is available from Professor Kou (`sk75@columbia.edu`).

neutral distribution, leading to differences in option prices, particularly for deep in- and out-of-money options, resulting in a skewed implied volatility surface.[4]

The estimated DEJD parameters reported below indicate strong negative skewness in both the risk neutral and the physical returns distribution, suggesting that the probability of a large decrease in stock prices exceeds the probability of a large increase. Jackwerth and Rubinstein (1996) termed this phenomenon as "crashophobia". The economic rationale for crashophobia is that put options are used as hedging instruments to protect against large downward movements in stock prices. This demand by investors due to portfolio insurance strategies has increased the price of protection (resulting in a "crash premium") and therefore the left tail of the risk neutral distribution has more weight.

3 Maximum likelihood estimation

We rely on MLE method to obtain parameter estimates for DEJD because it has desirable statistical properties. Other methods for the estimation of JD processes, including the generalized method of moments, the simulated moment estimation, and MCMC methods, among others are surveyed in Aït-Sahalia and Hansen (2004). The details on MLE estimation of jump-diffusion processes can be found in Sorensen (1991), who proves that for large samples, MLE is the best method of estimation, because under mild regularity conditions, the estimated parameter are consistent, asymptotically normal and asymptotically efficient; also see Bates (2003b) and the reference section of Aït-Sahalia (2002). However, MLE requires a complete specification of the transition density, which for nonlinear models may be difficult to obtain. Fortunately, the DEJD is a linear process with independent increments and an explicit transition density. Moreover, the selected distributions for the jump components have properties that make the MLE tractable. Lastly, Aït-Sahalia (2004) has shown that MLE offers advantages in disentangling jumps from diffusion.

Let $D = \{S(0), S(1), S(2), \ldots, S(M)\}$ denote the the realizations of stock price at equally-spaced times $k = 0, 1, 2, \ldots, M$. The one period rate of return $r_i = \ln S(i) - \ln S(i-1)$ is IID. As shown in Appendix 1, the unconditional density of $s = 1$ period returns, $f(r)$, is:

$$f(r) = \mathrm{e}^{-(\lambda_u+\lambda_d)} f_{0,0}(r) + \mathrm{e}^{-\lambda_u} \sum_{n=1}^{\infty} P(n, \lambda_d) f_{0,n}(r) + \mathrm{e}^{-\lambda_d} \sum_{m=1}^{\infty} P(m, \lambda_u) f_{m,0}(r) + \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} P(n, \lambda_d) P(m, \lambda_u) f_{n,m}(r)$$

[4] Assuming a range of parameter values, Kou (2002) and others provide evidence that the DEJD can generate the widely documented patterns of volatility smile and smirk. Using Kou's option pricing formula, firm specific data, and the parameter estimates reported here, we arrived at identical conclusions. These findings will not be presented to save space but are available in Ramezani and Zeng (2005).

where $P(n,\lambda) = \frac{e^{-\lambda}\lambda^n}{n!}$ and $f_{n,m}(r)$ ($n \geq 0$ and $m \geq 0$) is the conditional density for one period returns, conditional on the given numbers of up and down jumps (m,n). The log-likelihood given M equally spaced returns observations is:

$$L(D;\lambda_u,\lambda_d,\eta_u,\eta_d,\mu,\sigma^2) = \sum_{i=1}^{M} \ln(f(r_i)) \tag{5}$$

The unconditional density, $f(r)$, is a mixture density (i.e., a Poisson weighted sum of four conditional densities). Kiefer (1978) and Honoré (1998) have shown that for mixture densities, care must be taken to ensure that the log-likelihood function remains bounded by properly restricting the admissible parameter space. Otherwise a singularity problem arises, and the log-likelihood function becomes infinite. For the DEJD, so long as σ is bounded away from zero, the likelihood function will not explode and the MLE obtained is consistent and asymptotically normal. Hamilton (1994, page 689) and Kiefer (1978) have offered remedies to deal with this problem. As Hamilton (1994) shows singularities do not pose a major problem so long as the selected numerical maximization procedure converges to a local maxima. Moreover, standard errors for the estimates can be constructed using the information matrix (see below).

The Newton–Raphson method has been a widely used numerical procedure for likelihood optimization. This method requires the first and second order derivatives of the log-likelihood function. Such derivatives are difficult to compute for the DEJD model. To avoid this difficulty we use Powell's method. The latter method is a direction set optimization procedure that produces a set of mutually conjugate directions. Then it iteratively applies the line optimization to each conjugate direction to obtain the optimum. The Powell's method is a quadratically convergent algorithm that does not necessitate the use of the derivatives. Hamilton (1994, page 139) provides a complete description of Powell's optimization procedure. The optimization programs we use (subroutine "POWELL" and "QGAUS") are taken from Press et al. (1992).

The likelihood function in (5) involves double infinite summations and double improper integrals. First, piecewise Gaussian quadratures are employed to compute the integrals (the subroutine "QGAUS" is used). We find that for plausible parameter values, lower bound truncation of the integrals at (−2.0) provides six digit accuracy for most cases. Next, the infinite sums are calculated using the usual termination criterion; if $S_n = \sum_{i=1}^{n} X_i$, then we stop the summation if $2|X_{n+1}| \leq \text{FTOL} \times (|S_n| + |S_{n+1}|)$. We choose $\text{FTOL} = 10^{-10}$, which guarantees at least eight digits accuracy. Standard error for the estimates is obtained by the *outer-product* method, which is based on the first derivative of the likelihood function [see Hamilton (1994), page 143]. We performed extensive Monte Carlo simulation and found that our programs and the estimation procedures are accurate and reliable. In particular, starting with a set of assumed parameter values, we simulated data from DEJD specification and applied the Fortran programs to estimate the parameters from the simulated data. The resulting estimates are close to the assumed values and are always within two standard error bounds of the estimates. The Monte Carlo results will not be reported here to save space.

4 Model selection

The DEJD is compared with the LJD, and the GBM. For model selection, we adopt the widely-used Bayesian information criterion (BIC) proposed by Schwarz (1978). Unlike significance tests, BIC allows comparison of more than two models at the same time and does not require that the alternatives to be nested. BIC is a "conservative" criterion (relative to AIC) in the sense that it heavily penalizes over parametrization.

Suppose the kth model M_k, has parameter vector θ_k, where θ_k consists of n_k independent parameters to be estimated. Denote $\hat{\theta}_k$ as the MLE of θ_k. Then, BIC for Model M_k is defined as:

$$BIC_k = -2 \log f(D|\hat{\theta}_k, M_k) + n_k \log(m),$$

where m is the number of observations in data set D and $f(D|\hat{\theta}_k, M_k)$ is the maximized likelihood function. Clearly the best "fit" model is one with the smallest BIC. We provide comparisons of the DEJD ($\theta_{\text{DEJD}} = (\mu, \sigma, \lambda, p, \eta_u, \eta_d)$) relative to LJD ($\theta_{\text{LJD}} = (\lambda, \alpha, \beta, \mu, \sigma)$) and the GBM ($\theta_{\text{GMB}} = (\mu, \sigma)$).

5 Data and results

The estimation of DEJD, as well as the calculation of the standard errors, is computationally time consuming because the likelihood function involves double infinite summations and double improper integrals. On a high-performance computer (Itanium II, 1.3 GHz CPU with 1 GB RAM using HP-Fortran), 4–7 h of computation time is needed to obtain the first set of parameter estimates for a series consisting of 1,256 observations. To ensure the likelihood is fully optimized, at least two rounds of estimation are conducted for each series. While "good" initial parameter values may be helpful, and in some cases such values were obtained by cumulant matching methods, they are not necessary.

Another challenge arising during the estimation process is the possibility that the likelihood function may explode rather than converge. To avoid this singularity problem, we choose a range of initial values to ensure the parameter space is large enough to cover the true parameter values. We also ensure that the likelihood function obtained by Powell's method converges. Hence the conditions described in Hamilton (1994) and Kiefer (1978) are met and the consistency and asymptotic normality of the obtained maximum likelihood estimates are guaranteed.

To permit comparison with other studies, we focus on daily returns for the S&P-500 and the NASDAQ composite indexes. We also use daily returns for ten individual stocks, five with low kurtosis (range of 3–6) and five with high kurtosis (greater than 10).[5] The selected firms, which trade on NASDAQ (5) and NYSE (5), are followed by a large number of analysts, and are highly liquid. These characteristics are important given the event driven nature of the JD models. We estimate GBM, LJD, and DEJD for each series and use the BIC criterion to identify firms and indexes for which the DEJD specification fits the data best.

We use daily (value weighted) returns for the S&P-500 and the NASDAQ composite indexes. Two S&P-500 return series are utilized: SPD1 (SPD2) is the daily raw (dividend adjusted) returns for the period 7/1962 through 12/2003 ($N = 10,446$). The NASDAQ series, NASD, spans the period 1/1973 through 12/2003 ($N = 7,828$). No dividend adjusted series are available for NASDAQ index since few firms on this exchange pay dividends. Finally, the data on individual stocks spans the period 1/1999 through 12/2003 ($N = 1,256$).

Table 2 presents the sample statistics for the indexes and individual stocks returns. The large range of return values, particularly for the indexes, reflect significant booms and crashes that occurred during the sample period. All returns are highly skewed and exhibit large tail probabilities. Note that the skewness and kurtosis of individual stocks and the indexes are very similar. The last three columns of Table 2 report the number of days in the sample with positive returns (Up Freq), no change in returns, and negative returns (Down Freq). As expected, there are fewer days with no change in prices and comparable number of days with up and down price movements.

Table 3 reports the BIC values for the three alternative specifications. As noted above, the model with the smallest BIC provides the best fit to the data. The results in the table can be summarized as follows: The DEJD is clearly superior to GBM and LJD for both the S&P-500 and NASDAQ series. For individual stocks, we find that DEJD provides the best fit for seven out of ten series. Among high (low) kurtosis series, DEJD provides the best fit in four out of five (three out of five) instances. Consistent with previous studies, the GBM fails to beat the jump diffusion specifications for all returns series.

Parameter estimates and their standard errors for GBM, LJD, and DEJD are presented in Table 4. Estimates that are significant at 95% or higher confidence level appear in bold face. In discussing these results, we focus on the raw daily S&P-500 returns (SPD1, first row of the table), which is widely used in other empirical studies. Considering the LJD parameters, it appears that a jump in return process occurs approximately once every 24 days (λ^{-1}). The average jump size, in daily percentage form, is 0.08%, though this parameter is statistically insignificant. The standard deviation of daily jump magnitudes is 2.37%. As expected, under the JD specifications, the estimated mean and volatility associated with the continuous component of the process are smaller. Using the same daily return series for the period 1980–1996, Andersen et al. (2002) report estimates for GBM, LJD, and a model with stochastic volatility and jumps. Our parameter estimates for the GBM and LJD are comparable to

[5] We thank the referee for suggesting this approach in selecting individual stocks. An expanded version of this study containing results for 100 stocks is available from the authors upon request.

Table 2 Sample statistics for S&P500, NASDAQ, and 10 individual stocks

Name	Ticker	Minimum	Median	Maximum	Mean	SD	Skewness	Kurtosis	Up Freq	No Change	Down Freq
S&P-500	SPD1	−0.2047	0.0004	0.0910	0.0003	0.0095	−0.9448	25.758	5,464	45	4,937
S&P-500	SPD2	−0.1957	0.0005	0.0886	0.0005	0.0095	−0.8327	22.119	5,579	0	4,867
NASDAQ	NASD	−0.1132	0.0011	0.1427	0.0005	0.0124	−0.0823	10.778	4,415	0	3,413
ATT CP	T	−0.1908	−0.0022	0.2317	−0.0006	0.0296	0.4847	5.732	560	20	676
CYPRESS BIO	CYPB	−0.2941	0.0000	0.7708	0.0021	0.0724	1.8645	13.889	516	120	620
HYCOR BIO	HYBD	−0.4590	0.0000	0.5852	0.0033	0.0680	1.3077	10.906	541	125	590
INTEL CP	INTC	−0.2203	−0.0004	0.2012	0.0007	0.0360	−0.0916	3.120	621	5	630
LIFECORE BIO	LCMB	−0.5754	0.0000	0.5387	0.0009	0.0499	−0.1043	30.443	538	111	607
MFRI CP	MFRI	−0.2264	0.0000	0.5044	0.0009	0.0534	1.4315	12.678	509	262	485
MONSANTO	MON	−0.1514	0.0000	0.1260	0.0007	0.0267	0.1209	3.042	615	34	607
TIMBERLAND	TBL	−0.1455	0.0000	0.2004	0.0017	0.0298	0.5169	4.323	625	20	611
TIFFANY	TIF	−0.2084	−0.0002	0.2298	0.0015	0.0324	0.4135	4.919	615	12	629
XICOR	XICO	−0.6361	0.0000	0.5514	0.0039	0.0658	0.3423	11.830	595	54	607

Raw and dividend adjusted daily returns for S&P500 (SPD1 and SPD2) span the period 7/1962 through 12/2003 ($N = 10,446$). Daily NASDAQ returns, NASD ($N = 7,828$), span the period 1/1973 through 12/2003. The daily returns for the individual stocks span the period 1/1999 through 12/2003 ($N = 1,256$). The number of days in the sample with positive returns (Up freq), no change in returns, and negative returns (Down freq) appear in the last columns

Table 3 Comparison of alternative models

Ticker	GBM	LJD	DEJD
SPD1	−67, 724.75	−69, 176.61	**−69, 599.32**
SPD2	−67, 683.35	−69, 194.25	**−69, 521.58**
NASD	−46, 530.49	−48, 698.85	**−49, 402.71**
T	−5, 261.65	−5, 405.34	**−5, 406.17**
CYPB	−3, 018.41	**−3, 409.86**	−3, 408.85
HYBD	−3, 176.25	−3, 557.46	**−3, 564.59**
INTC	−4, 769.75	−4, 833.66	**−4, 835.93**
LCMB	−3, 951.71	−4, 526.84	**−4, 544.81**
MFRI	−3, 782.55	−4, 302.66	**−4, 476.41**
MON	−5, 523.13	−5, 634.04	**−5, 640.11**
TBL	−5, 247.67	**−5, 393.48**	−5, 392.69
TIF	−5, 035.51	**−5, 172.60**	−5, 172.03
XICO	−3, 257.33	−3, 481.52	**−3, 487.33**

The table contains the BIC values for three alternative specifications. Bold face shows the minimum BIC value and indicates that the specification fits the data better than the alternatives

those reported by these authors. For example their EMM estimates for GBM are $\mu = 0.0004, \sigma = 0.0059$; and for LJD, $\mu = 0.0003, \sigma = 0.0062, \lambda = 0.051$, $\alpha = 0$ (fixed), and $\beta = 0.0104$ (page 1265).[6]

Turning to the DEJD, all parameter estimates for S&P-500 and NASDAQ are statistically significant. Returning to SPD1, it appears that good and bad news arrive, roughly once every 2 days ($\lambda_u^{-1} = 2.15$ and $\lambda_d^{-1} = 1.78$, respectively). The mean up and down jump magnitudes, in daily percentage form, are 0.57 and 0.54%, respectively, (η_u^{-1} and η_d^{-1}). The estimated parameters for the dividend adjusted series (SDP2) are nearly identical. The distinction between good and bad news seems more pronounced for the NASDAQ series. It appears that good and bad news arrive every 4 and 2 days, respectively, and the mean up and down jump magnitudes, in daily percentage form, are 1.04 and 0.92% respectively.

Our estimates of the intensities for S&P-500 and NASDAQ are high relative to other studies. The fact that our estimates of λ parameters are so high seems to imply that high-frequency jumps are needed to fit the index return data better (even though technically the DEJD is not a high-frequency Lévy model). This implication is consistent with the finding of Huang and Wu (2004) that a high-frequency jump component in the return process is needed to fit the S&P-500 index better. Note that even though the Huang and Wu (2004) study is based on options, the implication for the importance of high-frequency jumps applies to the physical measure also as a high-frequency jump is of high-frequency regardless of the measure used.

The results for individual stocks in Table 4 appear somewhat mixed in that, unlike the indexes, the six parameters of DEJD are only significant for two stocks. For the remaining stocks, we find that the estimates of λ_u and η_u are

[6] Also see Andersen et al. (2002) table III on page 1249.

Table 4 Maximum Likelihood Estimates for GBM, LJD, and DEJD

Ticker	The GBM		Lognormal jump-diffusion (LJD)					Double exponential jump-diffusion (DEJD)					
	μ	σ	λ	α	β	μ	σ	λ_u	λ_d	η_u	η_d	μ	σ
SPD1	**0.0004**	**0.0094**	**0.0422**	0.0008	**0.0237**	**0.0003**	**0.0085**	**0.4640**	**0.5624**	**174.09**	**185.92**	**0.0007**	**0.0047**
	0.0001	0.0001	0.0103	0.0014	0.0008	0.0001	0.0001	0.0714	0.0933	0.43	0.44	0.0000	0.0000
SPD2	**0.0005**	**0.0095**	**0.0737**	0.0005	**0.0199**	**0.0004**	**0.0080**	**0.4719**	**0.5719**	**173.91**	**185.98**	**0.0008**	**0.0046**
	0.0001	0.0001	0.0106	0.0009	0.0005	0.0001	0.0001	0.0711	0.0922	0.36	0.38	0.0000	0.0000
NASD	**0.0056**	**0.0124**	**0.5522**	−0.005	**0.0137**	**0.0041**	**0.0057**	**0.2303**	**0.4368**	**95.90**	**110.38**	**0.0021**	**0.0050**
	0.0014	0.0001	0.0196	0.0003	0.0001	0.0001	0.0001	0.0238	0.0352	0.60	0.70	0.0000	0.0000
T	0.0001	**0.0296**	**0.1611**	**0.0133**	**0.0461**	**−0.0025**	**0.0222**	**0.3475**	0.0494	**47.35**	**28.97**	**−0.0059**	**0.0210**
	0.0008	0.0006	0.0329	0.0044	0.0028	0.0008	0.0006	0.1199	0.1225	2.42	2.15	0.0011	0.0000
CYPB	**0.0047**	**0.0724**	**0.4467**	**0.0253**	**0.0872**	**−0.0083**	**0.0338**	**0.4338**	0.3535	**17.08**	**25.75**	**−0.0092**	**0.0309**
	0.0020	0.0014	0.0423	0.0046	0.0035	0.0016	0.0006	0.1638	0.1972	1.87	1.82	0.0022	0.0000
HYBD	**0.0055**	**0.0679**	**0.6761**	**0.0122**	**0.0710**	**−0.0046**	**0.0256**	**0.4660**	**0.7224**	**19.41**	**33.25**	**0.0013**	**0.0230**
	0.0019	0.0014	0.0517	0.0029	0.0023	0.0014	0.0012	0.2365	0.3368	0.42	0.41	0.0003	0.0000
INTC	0.0014	**0.0360**	**0.0527**	0.0020	**0.0754**	0.0011	**0.0323**	**0.3390**	0.0610	**47.22**	**24.49**	−0.0036	**0.0281**
	0.0010	0.0007	0.0298	0.0133	0.0094	0.0010	0.0008	0.1590	0.1541	1.88	1.69	0.0016	0.0000
LCBM	0.0002	**0.0499**	**0.2156**	**0.0160**	**0.0824**	**−0.0022**	**0.0261**	**0.3015**	0.1261	**24.22**	**20.57**	**−0.0051**	**0.0234**
	0.0001	0.0010	0.0340	0.0053	0.0031	0.0009	0.0007	0.0613	0.0735	1.77	1.56	0.0010	0.0000
MFRI	0.0023	**0.0534**	**0.4827**	0.0013	**0.0676**	0.0005	**0.0206**	**0.7179**	**0.7517**	**33.07**	**35.67**	**0.0002**	**0.0038**
	0.0015	0.0011	0.0434	0.0032	0.0024	0.0010	0.0008	0.1556	0.1882	0.11	0.15	0.0000	0.0000
MON	0.0011	**0.0267**	**0.5510**	**0.0036**	**0.0287**	−0.0011	**0.0156**	**0.5369**	0.2925	**61.35**	**56.18**	−0.0027	**0.0155**
	0.0007	0.0005	0.0483	0.0017	0.0013	0.0085	0.0007	0.2537	0.2740	5.35	4.76	0.0015	0.0000
TBL	**0.0021**	**0.0298**	**0.3171**	**0.0080**	**0.0383**	−0.0007	**0.0197**	**0.5031**	0.3218	**51.70**	**56.52**	−0.0023	**0.0172**
	0.0008	0.0006	0.0383	0.0028	0.0020	0.0009	0.0007	0.2335	0.2637	4.31	3.94	0.0015	0.0000
TIF	**0.0020**	**0.0324**	**0.2342**	**0.0142**	**0.0437**	−0.0015	**0.0232**	**0.2986**	0.0330	**39.95**	**24.61**	**−0.0043**	**0.0238**
	0.0009	0.0006	0.0354	0.0037	0.0024	0.0009	0.0007	0.1336	0.1336	1.94	1.79	0.0012	0.0000
XICO	**0.0060**	**0.0658**	**0.2247**	**0.0398**	**0.0891**	**−0.0040**	**0.0446**	**0.3065**	0.0132	**18.60**	**8.37**	**−0.0010**	**0.0456**
	0.0019	0.0013	0.0350	0.0072	0.0039	0.0017	0.0013	0.0590	0.0473	0.92	1.30	0.0021	0.0000

Standard Errors appear below the estimates. Parameter estimates that are significant at 95% or higher confidence level appear in bold face. The raw and dividend adjusted daily returns for S&P500 (SPD1 and SPD2) span the period 7/1962 through 12/2003 ($N = 10,446$). Daily NASDAQ returns (NASD) span the period 1/1973 through 12/2003 ($N = 7,828$). Dividend and split adjusted daily returns for the 10 stocks span the period 1/1999 through 12/2003 ($N = 1,256$)

always significant. However, the estimates of λ_d are insignificant even though the estimates of η_d are always significant. This is puzzling as it appears that the combined effect of news arrival and jumps is more pronounced (in a statistical sense) for upward rather than the downward price moves.

Considering the results in Table 4, particularly for individual stocks, it can be argued that the distributional choice for the jump magnitudes may be inadequate. To see this conjecture, note that in the presence of large amount of small-sized jumps, LJD and DEJD are expected to perform similarly. Indeed, absent a high number of large negative and positive jumps, it may be difficult to distinguish between the two alternatives. This is because the exponential distribution, with it probability mass centered at the origin, predicts a rather large number of smaller jumps, which is consistent for the reported results for the indexes.

The estimate of η_u in Table 4 ranges from 17 to 174, with the majority exceeding 40. Given these parameter values, the density function for the up jumps, $f_{V^u}(x)=(\frac{\eta_u}{x^{\eta_u+1}}),\ \ V^u \geq 1$, appears to peak at 1 (i.e., a price jump of 0%), and drops sharply as x increases. For example, if $\eta_u = 80$, then over 95% of the up jumps will be less than 3% in magnitude. For individual stocks, price jumps of less than 3% probably should be regarded as normal part of the GBM. Therefore, for most price increases, the DEJD specification may have a difficult time separating normal price movements from jumps. Perhaps a better choice of the distribution should have a single mode above 3%. Given this observation, future research should consider other distributions such as Gamma. A better distributional choice will likely enhance the fit, though the likelihood function is bound to become more complicated and the analytical advantages of DEJD for pricing options may disappear. The foregoing discussion also strengthens the case calling for high-frequency Lévy model, as pointed out by Huang and Wu (2004). This is because a high-frequency jump structure allows for both high-frequency small jumps and rare large jumps.

Overall, the parameter estimates in Table 4 have reasonable values and are informative. In particular, the table shows that for *equity returns*, the symmetric version of DEJD (i.e., $\eta_u = \eta_d = \eta$ and $\lambda_u = \lambda_d = \lambda$), rarely occurs. As noted earlier, several important applications of the DEJD invoke this assumption. The data in Table 4 provides no empirical support for this assumption.

Moreover, considering the DEJD parameters in Table 4, the intensity of news arrival appears large, while the mean jump amplitudes seem relatively small. It is important to note that it is this combination of high jump intensity and small jump magnitudes that generates the high peak and the leptokurtic features of return distribution under DEJD. Finally, the addition of a jump component significantly changes the estimated drift and volatility parameters associated with the continuous part of the process, as expected.

The parameter estimates in Table 4 in conjunction with the formula in Table 1 can be used to calculate the moments of the returns distribution. Comparison of the calculated moments with the sample moments, particularly skewness and kurtosis (not tabulated), shows that the DEJD matches the empirical moments

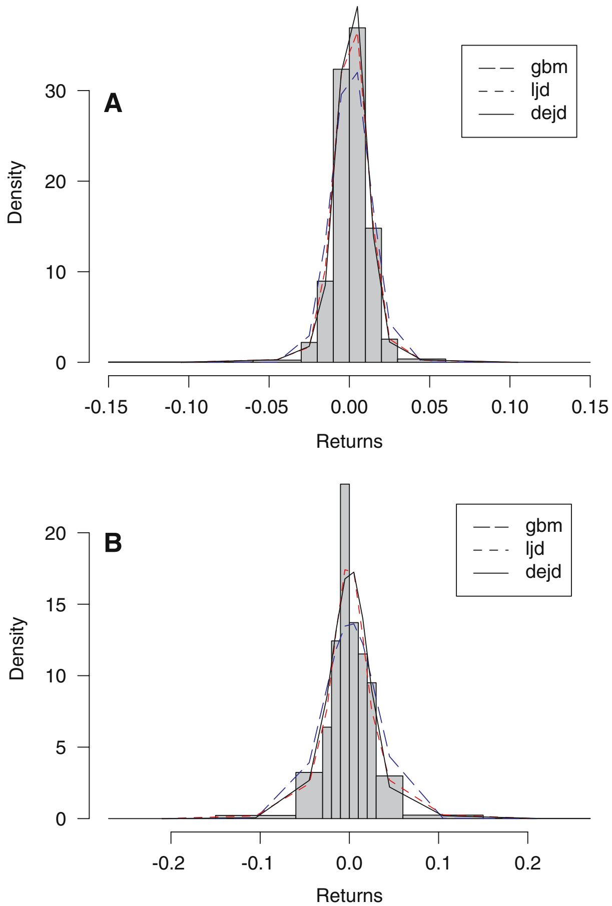

Fig. 1 **a** Fitted and actual returns distributions for SPD. **b** Fitted and actual returns distributions for TBL

quite well. Utilizing the parameter estimates in Table 4, we plot the fitted densities for GBM, LJD, and DEJD against the sample histograms for two series. Figure 1a shows the histogram and the fitted distributions for the daily S&P-500 (SPD1). The DEJD clearly better matches the high peak, skewness, and kurtosis of the returns. Similarly, Fig. 1b shows the histogram and the fitted distributions for Timberland Corporation (TBL). As shown in Table 4, for this stock, LJD provides a better fit than DEJD. While Fig. 1b appears to confirm this finding, it is clear that both LJD and DEJD fail to match the peak of the histogram, suggesting that other distributions may better fit TBL data.

6 Conclusions

The paper presented an empirical assessment of the DEJD. We provided maximum likelihood estimates for a sample of CRSP firms, the S&P-500, and the NASDAQ composite. Using the BIC criterion, we assessed the performance of DEJD relative to the LJD and GBM. We found that both DEJD and LJD are superior to GBM. Moreover, for majority of the individual stocks and the indexes, the DEJD provides a better fit than LJD.

These finding are significant because while the LJD is simpler from an econometric perspective (transition density has a convenient expression), the DEJD is better suited for pricing derivatives, particularly path-dependent contingent claims. Accordingly, it is important to establish which model is more compatible with indexes or individual equity data. Our results shows that for indexes, the DEJD performs better than LJD. For individual stocks, the DEJD does not always dominate the LJD. Assessing the performance of DEJD for a large number of individual stocks is a major undertaking and outside the scope of the present study. Such study requires extensive computational time. As noted above, we found it necessary to conduct at least two rounds of estimation to ensure maxima and the convergence of the likelihood function.

There are a number of other interesting directions for future extensions of this work. As a starting point, other estimation techniques, such as the generalized method of moments and its variants may be utilized. As Eraker et al. (2003) and others have shown, stochastic volatility is an important component of the return process and should be integrated into the DEJD specification. As an example, it may be interesting to nest ARCH effects into the DEJD model. This extension has already been proposed by Keppo et al. (2003) but it results in a very complicated likelihood function. With the moment based methods it may be simpler to determine whether stochastic volatility remains important when the jump component of return process has a more complex structure as in DEJD. Time-varying jump intensities, as proposed by Andersen et al. (2002), offer another direction to extend the DEJD specification.

Acknowledgments We thank an anonymous referee, David Bates, Zhiwu Chen, Rama Cont, Chin-Shan Chuang, Steven Kou, Thomas Kurtz, Artur Sepp, Luc Soenen and seminar participants at the University of Cambridge, Cornell University, the Federal Reserve Board, University of Missouri (Kansas City) and the University of Wisconsin (Madison and Milwaukee) for valuable comments and suggestions.

Appendix 1: The derivation of the density of returns

Let $N_s^u = m$ and $N_s^d = n$ be the number of up- and down-jumps during the time interval with length s. The conditional densities of s period returns can be derived under four combination of m and n: $m = 0$ and $n = 0$ (no jumps occur); $m = 0$ and $n \geq 1$ (only down-jumps occur); $m \geq 1$ and $n = 0$ (only up-jumps occur); and $m \geq 1$ and $n \geq 1$ (both types of jumps occur). All four conditional densities can be derived using convolution techniques and distributional properties. Before deriving the conditional density, we note some useful facts about Pareto, Beta and exponential distributions (Patel et al., 1976):

F1. If $V^u \sim$ Pareto (η_u), then $Y^u = \ln(V^u) \sim \exp(\eta_u) = \Gamma(1, \eta_u)$.
F2. If $V^d \sim$ Beta $(\eta_d, 1)$, then $-Y^d = -\ln(V^d) \sim \exp(\eta_d) = \Gamma(1, \eta_d)$.

F3. If $X = Y_1 + Y_2 + \cdots + Y_n$, where $Y_i \sim \exp(\theta)$ and they are independent, then $X \sim \Gamma(n,\theta)$.

Let $U = \sum_{i=1}^{N_s^u} Y_i^u > 0, D = \sum_{i=1}^{N_s^d} Y_i^d < 0$ and $T = U + D$. Then s period return can be written as $r(s) = (\mu - 0.5\sigma^2)s + Z(s) + U + D$. For $N_s^u = m \geq 1$ the conditional distribution of U (by F1 and F3) is $U|m \sim \Gamma(m, \eta_u)$ with the density:

$$f_{U|m}(U) = \frac{\eta_u^m}{(m-1)!} U^{m-1} \mathrm{e}^{-\eta_u U}$$

Similarly, for $N_s^d = n \geq 1$ the conditional distribution of D is $-D|n \sim \Gamma(n, \eta_d)$ with the density:

$$f_{D|n}(D) = \frac{\eta_d^n}{(n-1)!} (-D)^{n-1} \mathrm{e}^{\eta_d D}$$

Applying the above two results, the conditional density of $T = U + D$, given $m \geq 1$ and $n \geq 1$, is:

$$\begin{aligned} f_{T|m,n}(t) &= \int_{-\infty}^{\infty} f_D(x) f_U(t-x) \mathrm{d}x \\ &= \frac{\eta_u^m \eta_d^n \mathrm{e}^{-\eta_u t}}{(m-1)!(n-1)!} \int_{-\infty}^{0 \wedge t} (-x)^{n-1} (t-x)^{m-1} \mathrm{e}^{(\eta_u + \eta_d)x} \mathrm{d}x \qquad \text{(A1)} \end{aligned}$$

Now, we are ready to determine all four conditional densities. For the case $m = 0$ and $n = 0$, the conditional density is that of $N((\mu - 0.5\sigma^2)s, \sigma^2 s)$:

$$f_{r(s)|0,0}(r) = \frac{1}{\sqrt{2\pi s}\sigma} \mathrm{e}^{-\frac{1}{2\sigma^2 s}(r - \mu s + 0.5\sigma^2 s)^2} \qquad \text{(A2)}$$

When $m = 0$ and $n \geq 1$, the conditional distribution is the independent sum of $-\Gamma(n, \eta_d)$ and $\mathrm{N}((\mu - \frac{1}{2}\sigma^2)s, \sigma^2 s)$:

$$f_{r(s)|0,n}(r) = \frac{\eta_d^n}{(n-1)!\sqrt{2\pi s}\sigma} \int_{-\infty}^{0} (-x)^{n-1} \mathrm{e}^{\eta_d x - \frac{1}{2\sigma^2 s}(r - x - \mu s + 0.5\sigma^2)^2 s} \mathrm{d}x \qquad \text{(A3)}$$

Similarly, for $m \geq 1$ and $n = 0$, the conditional distribution is the independent sum of $\Gamma(m, \eta_u)$ and $\mathrm{N}((\mu - \frac{1}{2}\sigma^2)s, \sigma^2 s)$:

$$f_{r(s)|m,0}(r) = \frac{\eta_u^m}{(m-1)!\sqrt{2\pi s}\sigma} \int_{0}^{\infty} (x)^{m-1} \mathrm{e}^{-\eta_u x - \frac{1}{2\sigma^2 s}(r - x - \mu s + 0.5\sigma^2 s)^2} \mathrm{d}x \qquad \text{(A4)}$$

Finally, for $m \geq 1$ and $n \geq 1$, the conditional distribution is the independent sum of the distribution for T and $\mathrm{N}((\mu - \frac{1}{2}\sigma^2)s, \sigma^2 s)$. Then the conditional density of $r(s)$ is:

$$f_{r(s)|m,n}(r) = \frac{\eta_u^m \eta_d^n}{(m-1)!(n-1)!\sqrt{2\pi s}\sigma} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{0\wedge t} (-x)^{n-1}(t-x)^{m-1} \mathrm{e}^{(\eta_u+\eta_d)x} \mathrm{d}x \right) \times \mathrm{e}^{-\eta_u t} \mathrm{e}^{-\frac{1}{2\sigma^2 s}(r-t-\mu s+0.5\sigma^2 s)^2} \mathrm{d}t \tag{A5}$$

Next we derive the unconditional density of $s = 1$ period returns. Letting $P(n, \lambda) = \frac{\mathrm{e}^{-\lambda}\lambda^n}{n!}$, the unconditional density for one period returns, $f(r)$, can be written as the Poisson weighted sum of the four conditional densities (A2–A5):

$$f(r) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} P(n,\lambda_d)P(m,\lambda_u)f_{n,m}(r) = \mathrm{e}^{-(\lambda_u+\lambda_d)} f_{0,0}(r) + \mathrm{e}^{-\lambda_u} \sum_{n=1}^{\infty} P(n,\lambda_d) f_{0,n}(r) + \mathrm{e}^{-\lambda_d} \sum_{m=1}^{\infty} P(m,\lambda_u) f_{m,0}(r) + \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} P(n,\lambda_d)P(m,\lambda_u)f_{n,m}(r) \tag{A6}$$

and as equation (A6) shows, the unconditional distribution of returns is a *mixture density*.

References

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

Aït-Sahalia, Y.: Disentangling diffusion from jumps. J Financ Econ **74**, 487–528 (2004)

Aït-Sahalia, Y., Hansen, L.P.: Handbook of Financial Econometrics. Amsterdam: North-Holland (2004)

Andersen, L., Andreasen, J.: Jump-diffusion processes: Volatility smile fitting and numerical methods for pricing. Rev Derivative Res **4**, 231–262 (2000)

Andersen, T.G., Benzoni, L., Lund, J.: An empirical investigation of continuous-time equity return models. J Finance **57**, 1239–1284 (2002)

Bates, D.S.: Empirical option pricing: a retrospection. J Econom **116**(1–2), 387–404 (2003a)

Bates, D.S.: Maximum likelihood estimation of latent affine processes. Working Paper, University of Iowa (2003b)

Broadie, M., Yamamoto, Y.: Application of the fast gauss transform to option pricing. Manage Sci **49**(8), 1071–1088 (2003)

Carr, P., Madan, D.: Option pricing using the fast fourier transform. J Comput Finance **2**(4), 61–73 (1999)

Carr, P., Wu, L.: Time-changed Lévy processes and option pricing. J Financ Econ **17**(1), 113–141 (2004)

Chernov, M., Gallant, A.R., Ghysels, E., Tauchen, G.: Alternative models for stock price dynamics. J Econom **116**(1–2), 225–257 (2003)

Cont, R., Tankov, P.: Financial modelling with jump processes. London: Chapman & Hall / CRC, Financial Mathematic Series (2003)

Cont, R., Tankov, P.: Nonparametric calibration of jump-diffusion option pricing models. J Comput Finance **7**(3), 1–49 (2004)

dĤalluin, Y., Forsyth, P., Vetzal, K.: Robust numerical methods for contingent claims under jump diffusion processes. Working Paper, School of Computer Science, University of Waterloo (2004)

Duffie, D., Pan, J., Singleton, K.J.: Transform analysis and asset pricing for affine jump-diffusions. Econometrica **68**(6), 1343–1376 (2000)

Eraker, B., Johannes, M., Polson, N.: The impact of jumps in volatility and returns. J Finance **58**(3), 1269–1300 (2003)

Garcia, R., Ghysels, E., Renault, E.: The econometrics of option pricing, Working Paper, Université de Montréal - CIREQ / Département de sciences économiques (2004)

Hamilton, J.D.: Time Series Analysis. Princeton: Princeton University Press (1994)

Honoré, P.: Pitfalls in estimating jump diffusion models, Working Paper, University of Aarhus (1998)

Huang, J., Huangm M.: How much of the corporate-treasury yield spread is due to credit risk?, Working Paper, Graduate School of Business, Stanford University (2003)

Huang, J., Wu, L.: Specification analysis of option pricing models based on time-changed Lévy processes. J Finance **59**(3), 1405–1439 (2004)

Jackwerth, J., Rubinstein, M.: Recovering probability distributions from option prices. J Finance **51**, 1611–1631 (1996)

Keppo, J., Meng, X., Shive, S., Sullivan, M.: Modelling and hedging options under stochastic pricing parameters, Working Paper, Industrial and Operations Engineering, University of Michigan at Ann Arbor (2003)

Kiefer, N.M.: Discrete parameter variation: efficient estimation of switching regression model. Econometrica **46**(2), 427–34 (1978)

Kou, S.: A jump diffusion model for option pricing. Manage Sci **48**(8), 1086–1101 (2002)

Kou, S., Wang, H.: Option pricing under a double exponential jump diffusion model. Manage Sci **50**(9), 1178–1192 (2004)

Lee, R.W.: Implied volatility: Statics, dynamics, and probabilistic interpretation. Recent advances in applied probability (2006), (inpress)

Lewis, A.: A simple option formula for general jump-diffusion and other exponential Lévy processes. Envision financial system and optioncitynet (2001)

Maheu, J.M., McCurdy, T.: News arrival, jump dynamics and volatility components for individual stock returns. J Finance **59**(2), 755–768 (2004)

Merton, RC.: Option pricing when underlying stock returns are discontinuous. J Financ Econ **3**, 224–244 (1976a)

Merton, R.C.: The impact on option pricing of specification error in the underlying stock price returns. J Finance, **31**(2), 333–350 (1976b)

Milgrom, P.R.: Good news and bad news: representation theorems and applications. Bell J Econ **12**, 380–391 (1981)

Naik, V.: Option valuation and hedging strategies with jumps in the volatility of asset returns. J Finance **48**(5), 1969–1984 (1993)

Patel. J., Kapadia, C.H., Owen, D.B.: Handbook of Statistical Distributions. New York: Marcel Dekker Inc. (1976)

Press, S.J.: A compound events model for security prices. J Busi **40**, 317–335 (1967)

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in FORTRAN. London: Cambridge University Press (1992)

Protter, P.: Stochastic Integration and Differential Equations. Berlin Heidelberg New York: Springer (1991)

Ramezani, C., Zeng, Y.:Maximum likelihood estimation of asymmetric jump-diffusion processes: Application to security prices, Working Paper, Department of Mathematics and Statistics, University of Missouri, Kansas City, see http://papers.ssrn.com/sol3/papers.cfm?abstract_id=606361 (1998)

Ramezani, C., Zeng, Y.: An empirical assessment of the double exponential jump-diffusion process, Working Paper, Department of Mathematics and Statistics, University of Missouri, Kansas City, see http://papers.ssrn.com/sol3/papers.cfm?abstract_id=606101 (2005)

Schwarz, G.: Estimating the dimension of a model. Ann Stat **6**, 461–464 (1978)

Sepp, A.: Analytical pricing of double-barrier options under a double-exponential jump diffusion process: Applications of laplace transform. Int J Theor Appl Finance **7**(2), 151–175 (2004)
Sorensen, M.: Likelihood methods for diffusions with jumps. In: Prabhu, N.V., Basawa, I.V. (eds.) Statistical Inference in Stochastic Processes. New York: Marcel Dekker, Inc. (1991)
Tsay, R.S.: Analysis of Financial Time Series. Wiley Series in Probability and Statistics. Wiley Inter-Science (2002)