Contemporary Mathematics

Multinomial maximum likelihood estimation of market parameters for stock jump-diffusion models

Floyd B. Hanson, John J. Westman, and Zongwu Zhu

ABSTRACT. The estimated parameters of the log-return density for log-normal-diffusion, log-uniform jump process are found for an observed financial market distribution. When the observed data is collected into bins, it is shown that the appropriate parameter estimation method is the multinomial maximum likelihood estimation. This result is independent of the theoretical distribution, since it is only assumed that the observed distribution is the simulation of independent, identically distributed random variables. For the application to the theoretical jump-diffusion distribution, the estimation procedure is constrained by forcing the first two moments of the theoretical distribution to be the same as that for observed market distribution. The Standard and Poor's 500 stock index for the 1992-2001 decade is used as the observed market data. Numerically, the classical Nelder-Mead and our own direct search method are used to find the maximum likelihood estimation of the parameters. The results and performance of these numerical methods are compared along with the our older weighted least squares estimation method. The results of the two numerical approximations for the multinomial estimation methods were similar, but the weighted least squares results are not as good. In the severe test on the third and fourth moment measures, the multinomial based methods differed significantly from the same measures on the observed data, but did much better than the normal distribution based weighted least squares.

1. Introduction

While the log-normal diffusion or geometric Brownian motion stock-return model has been studied extensively and provides the basis for the Black-Scholes-Merton options model [**2, 10**], less is known about jump-diffusion stock-return models. Jump-diffusions provide added realism to the the stock return model since the jumps can include rare, large fluctuation to simulate stock market crashes or rallies. In addition, jumps allow higher moment features such as skewness and leptokurtic (peakedness) behavior in the stock log-return distribution, whereas for the

1991 *Mathematics Subject Classification.* Primary 60J75, 60J60; Secondary 91B28, 62M05, 91B82.

Key words and phrases. Jump-diffusions, multinomial maximum likelihood estimation, market parameters, computational finance.

The first author's work is supported by the National Science Foundation under Grant DMS-02-07081. The content of this material is that of the authors and does not necessarily reflect the views of the National Science Foundation.

normal distribution, as in the case of pure diffusion, the log-return is skewless and the coefficient of kurtosis has the normal neutral value of three.

Merton, in his pioneering discontinuous, jump-diffusion stock-return model [**10**, Chapter 9], used a Poisson process for the jump times and log-normally distributed jump-amplitudes for the jump process, i.e., a compound or space-time Poisson model. Kou [**9**] has developed a double exponential (gamma) distributed jump amplitude jump-diffusion model to include the negative skew, leptokurtic, and other properties in option pricing. Extensive probability properties are developed along with many special functions. Andersen, Benzoni and Lund [**1**] have made elaborate estimations to fit jump-diffusion models with stochastic volatility to the market. Their basic jump-amplitude part of the model is log-normal in various forms. Their estimation seems to be very robust, but the number of model parameters they estimate is unusually large so that the goodness of fit is not surprising. However, their justification of the need for including stochastic volatility in addition to jump processes is convincing.

There are several problems with these jump-diffusion models. One is that the jump-amplitude models are selected to produce exact analytic solutions rather than motivated by the market distributions. Crashes and rallies are relatively rare events so Poisson is a good model for the timing of jumps, but there is no good reason that the jump-amplitude should be log-normal or double exponential. The exponentially small tails of these distributions are not reasonable for modeling the thick tails that characterize the leptokurtic property of long time financial market distributions. Also, the log-normal and double exponential distributions typically peak in the center of the market distribution where the diffusion is the dominant part of the model, but at sufficiently small jump amplitudes a diffusion approximation of the space-time Poisson process should be valid. Another problem is that the doubly infinite range of the jump-amplitude distributions leads to severe restrictions on the optimal instantaneous stock fraction for optimal portfolios in terms of borrowing from and saving short-sales into a riskless asset [**6, 7**].

In our papers, most recently [**6, 7**], we have investigated the log-uniform distributed jump-amplitudes in the jump-diffusion model for fitting financial market distributions such as the Standard and Poor's 500 stock index. Since the uniform distribution has finite range, it lacks the severe restrictions of the infinite range models. Also, since the crashes and large rallies are rare, they are more like isolated outliers that are very difficult, if not impossible, to characterize. Hence, a uniform distribution is suitable by the principle of simplicity, at least until more information can be gleaned about these large deviations.

2. Log-Return for Log-Normal-Diffusion, Log-Uniform Jump Processes

2.1. Stock Return Process, $\mathbf{S(t)}$. Let $S(t)$ be the price of a stock or stock mutual fund at time t. Its dynamics evolve according to the following stochastic differential equation (SDE), with Itô's forward integration rule for complete specification,

$$dS(t) = S(t)[\mu_d dt + \sigma_d dW(t) + J(Q)dP(t)], \tag{2.1}$$

starting at $S(0) = S_0$, $S(0) > 0$, μ_d is the drift coefficient associated with the diffusion, σ_d is the diffusive volatility, $W(t)$ is the stochastic diffusion process, $J(Q)$ is the Poisson jump amplitude, Q is its underlying Poisson amplitude mark process whose relation to $J(Q)$ will be developed later, $P(t)$ is the standard Poisson jump process with joint mean and variance $\mathrm{E}[P(t)] = \lambda t = \mathrm{Var}[P(t)]$.

2.2. Stock Log-Return Process, $\mathbf{ln(S(t))}$. The stock log-return $\ln(S(t))$ can be transformed to a simpler jump-diffusion stochastic differential equation (SDE) upon use of the stochastic chain rule [**8**],

$$d[\ln(S(t))] = \mu_{ld}dt + \sigma_d dW(t) + \ln(J(Q)+1)dP(t), \tag{2.2}$$

where $\mu_{ld} \equiv \mu_d - 0.5\sigma_d^2$ can be called the log-diffusive (ld) drift.

2.3. Log-Uniform Jump Distribution for $\mathbf{Q = ln(J(Q)+1)}$. For simplicity, the mark to jump-amplitude relation is defined as $Q = \ln(J(Q)+1)$ for $J(Q) > -1$. Then, let the density of the jump amplitude mark Q be uniform

$$\phi_Q(q) = [H(Q_b - q) - H(Q_a - q)]/[Q_b - Q_a], \tag{2.3}$$

where $Q_a < 0 < Q_b$ and $H(x)$ is the Heaviside unit step function. The mark Q has moments, $\mu_j \equiv \mathrm{E}_Q[Q] = 0.5(Q_b + Q_a)$, $\sigma_j^2 \equiv \mathrm{Var}_Q[Q] = (Q_b - Q_a)^2/12$. The original jump-amplitude J has mean $\mathrm{E}[J(Q)] = (\exp(Q_b) - \exp(Q_a))/(Q_b - Q_a) - 1$ and log-uniform distribution

$$\Phi_J(x) = \ln((x+1)/(J_a+1))/\ln((J_b+1)/(J_a+1))$$

on $[J_a, J_b]$, where $J_a \equiv J(Q_a)$ and $J_b \equiv J(Q_b)$.

The finiteness of the density domain is important since infinite jump domains lead to unrealistic transaction restrictions. Jumps should be bounded if the market is bounded. Also, uniformness of the density is sufficient, since jumps are rare, outlier events, so that it is difficult to extract information to suggest any more complexity in the density. In addition, the tails are thick but not exponentially small like those of exponential or normal densities, so those densities are not appropriate for the jump component. As for the rest of the distribution between the tails, the diffusion dominates in the interior of the distribution and the diffusion approximation is valid for small jumps anyway, so the diffusion process should be adequate there.

2.3.1. *Basic Moments for Log-Return Increments $\Delta[\mathbf{ln(S(t))}]$ in Trading Time Increments $\mathbf{\Delta t}$.* The mean log-return is given by

$$M_1^{(\mathrm{jd})} \equiv \mathrm{E}[\Delta[\ln(S(t))]] = [\mu_{ld} + \lambda\mu_j]\Delta t, \tag{2.4}$$

and the log-return variance is

$$M_2^{(\mathrm{jd})} \equiv \mathrm{Var}[\Delta[\ln(S(t))]] = [\sigma_d^2 + \lambda(\sigma_j^2(1+\lambda\Delta t) + \mu_j^2)]\Delta t, \tag{2.5}$$

but if $\lambda\Delta t$ is sufficiently small Δt then the $\mathrm{O}^2(\lambda\Delta t)$ term can be omitted (for roughly 250 trading days per year, $\Delta t \simeq 4.0\text{e-}3$ years).

THEOREM 2.1. *: Probability density for Log-Normal-Diffusion, Log-Uniform-Jump-Amplitude Log-Return Increment $\Delta[\ln(S(t))]$ is given asymptotically by*

$$\begin{aligned} \phi^{(jd)}(x) \quad \sim \quad & p_0(\lambda\Delta t)\phi^{(n)}(x;\mu_{ld}\Delta t, \sigma_d^2\Delta t) \\ & + p_1(\lambda\Delta t)[\Phi^{(n)}(Q_b - x + \mu_{ld}\Delta t; 0, \sigma_d^2\Delta t) \\ & - \Phi^{(n)}(Q_a - x + \mu_{ld}\Delta t; 0, \sigma_d^2\Delta t)]/[Q_b - Q_a]\,, \end{aligned} \tag{2.6}$$

for sufficiently small Δt ($\Delta t \ll 1$) and $-\infty < x < +\infty$, where $p_k(\Lambda) = e^{-\Lambda}\Lambda^k/k!$ is the Poisson distribution with parameter Λ and k jumps. The superscript (n) simply *denotes normally distributed, so that $\phi^{(n)}(x;\mu,\sigma^2)$ is the normal density with mean μ and variance σ^2, while $\Phi^{(n)}(x;\mu,\sigma^2)$ is the corresponding normal distribution.*

SKETCH OF PROOF. The proof follows from the probability density of a triad [**6**], $\xi+\eta\zeta$, of independent log-return random variables: the normally distributed $\xi=\mu_{ld}dt+\sigma_d W(t)$, the uniformly distributed $\eta=Q$ and the Poisson distributed $\zeta=dP(t)$. The density of the sum $\xi+(\eta\zeta)$ is derived from the usual sum convolution theorem, but the density of product, $\eta\cdot\zeta$, is not well known (but treated by Hanson and Westman, [**5**]) and obeys the law of total probability [**14, 8**] for Poisson processes, or asymptotically

$$\phi_{\eta\zeta}(x)\sim p_0(\lambda\Delta t)\delta(x)+p_1(\lambda\Delta t)\phi_Q(x)\ ,$$

for $\Delta t\ll 1$, where λ is the jump rate, $\delta(x)$ is the Dirac delta function and $\phi_Q(x)$ is the jump amplitude mark density. □

3. Jump-Diffusion Parameter Estimation

The basic point of view, here, is that the financial markets are considered to be a moderate size simulation of a jump-diffusion process. However, other factors such as stochastic volatility may be needed to refine the jump-diffusion approximation.

3.1. Empirical Data. Standard and Poor's 500 (S&P500) stock index in the decade 1992-2001 [**15**] is viewed as one big mutual fund so that it is less dependent on the peculiar behavior of any one stock. Let $n^{(\mathrm{sp})}=2522$ be the number of daily closings $S_s^{(\mathrm{sp})}$ for $s=1:n^{(\mathrm{sp})}$, such that there are $ns=2521$ log-returns,

$$\Delta\left[\ln\left(S_s^{(\mathrm{sp})}\right)\right]\equiv\ln\left(S_{s+1}^{(\mathrm{sp})}\right)-\ln\left(S_s^{(\mathrm{sp})}\right)\ , \tag{3.1}$$

for $s=1:ns$ log-returns, with

- Mean:

$$M_1^{(\mathrm{sp})}=\frac{1}{ns}\sum_{s=1}^{ns}\Delta\left[\ln\left(S_s^{(\mathrm{sp})}\right)\right]\simeq 4.015\text{e-}4\ .$$

- Variance:

$$M_2^{(\mathrm{sp})}=\frac{1}{ns-1}\sum_{s=1}^{ns}\left(\Delta\left[\ln\left(S_s^{(\mathrm{sp})}\right)\right]-M_1^{(\mathrm{sp})}\right)^2\simeq 9.874\text{e-}5\ .$$

- Skewness coefficient: $\beta_3^{(\mathrm{sp})}\equiv M_3^{(\mathrm{sp})}/\left(M_2^{(\mathrm{sp})}\right)^{1.5}\simeq -0.2913<0$, where $\beta_3^{(n)}=0$ is the normal distribution value and $M_3^{(\mathrm{sp})}$ is the 3rd central log-return moment of the data.
- Kurtosis coefficient: $\beta_4^{(\mathrm{sp})}\equiv M_4^{(\mathrm{sp})}/\left(M_2^{(\mathrm{sp})}\right)^2\simeq 7.804>3$, where $\beta_4^{(n)}=3$ is the normal distribution value and $M_4^{(\mathrm{sp})}$ is the 4th central log-return moment of the data.

3.2. Multinomial Distribution of Simulation Frequencies. Data is further sorted into nb bins and in each centered bth bin $B_b=[\xi_{b-0.5\Delta\xi},\xi_{b+0.5\Delta\xi})$ for $b=1:nb$:

- Experimental S&P500 frequency for bin B_b: $f_b^{(\mathrm{sp})}$,

- Theoretical jump-diffusion frequency with parameter vector $\boldsymbol{x}$:

$$f_b^{(\mathrm{jd})}(\boldsymbol{x}) \equiv ns \int_{B_b} \phi^{(\mathrm{jd})}(\eta; \boldsymbol{x}) d\eta$$

  and the corresponding theoretical bin probability:

$$p_b = p_b^{(\mathrm{jd})}(\boldsymbol{x}) = f_b^{(\mathrm{jd})}(\boldsymbol{x})/ns = \int_{B_b} \phi^{(\mathrm{jd})}(\eta; \boldsymbol{x}) d\eta \ ,$$

- Simulated jump-diffusion frequency:

$$f_b^{(\mathrm{sim})} \equiv \sum_{s=1}^{ns} U\left(\Delta\left[\ln\left(S_s^{(\mathrm{sim})}\right)\right]; B_b\right)$$

for each bth bin for all $b = 1 : nb$ bins, where ns is the simulation sample size and $U(\eta; B_b)$ is the unit step or index function for η on set B_b.

The view here is that the underlying market distribution is a log-normal, log-uniform jump-diffusion distribution, so the bth simulation frequency

$$f_b^{(\mathrm{sim})} = f_b^{(\mathrm{sp})} \tag{3.2}$$

representation will be assumed in theory, but approximate in practice. The jump-diffusion distribution will serve as the theoretical distribution, so

$$f_b^{(\mathrm{th})} = f_b^{(\mathrm{jd})}(\boldsymbol{x}) \ . \tag{3.3}$$

However, the following theorem is quite general and only depends on the theoretical general bin frequencies $f_b^{(\mathrm{th})}$ and a corresponding independent identically distributed (IID) simulation with frequencies $f_b^{(\mathrm{sim})}$.

THEOREM 3.1. *Let $f_b^{(sim)}$ be the independent identically distributed bin frequency for $b = 1 : nb$ bins, $nb > 2$, and sample size ns simulations based upon the theoretical stochastic process (th) with probability density $\phi^{(th)}(\xi)$ with theoretical bin frequency $f_b^{(th)}$, then the simulation distribution is multinomial:*

$$\Phi^{(sim)}(\boldsymbol{k}) = Prob[\boldsymbol{f}^{(sim)} = \boldsymbol{k}] = ns! \prod_{b=1}^{nb} \frac{\left(f_b^{(th)}/ns\right)^{k_b}}{k_b!} \ , \tag{3.4}$$

where $\boldsymbol{k} = [k_b]_{nb\times 1}$, $\sum_{b=1}^{nb} k_b = ns$, $\boldsymbol{f} = [f_b]_{nb\times 1}$ and $\sum_{b=1}^{nb} f_b = ns$ are constraints, with bin mean

$$\mu_b^{(sim)} \equiv E\left[f_b^{(sim)}\right] = f_b^{(th)} = ns \cdot p_b^{(th)} = ns \cdot p_b \ ,$$

and bin variance

$$\left(\sigma_b^{(sim)}\right)^2 \equiv Var\left[f_b^{(sim)}\right] = f_b^{(th)} \cdot \left(1 - f_b^{(th)}/ns\right) \ ,$$

where $p_b = f_b^{(th)}/ns$ is the theoretical bin probability for $b = 1 : nb$ bins.

The single bin frequency is binomially distributed

$$\Phi^{(sim)}(k_b) \equiv E\left[\Phi^{(sim)}\left(\boldsymbol{f}^{(sim)}\right)\middle| f_b^{(sim)} = k_b\right] = ns! \frac{p_b^{k_b}(1-p_b)^{ns-k_b}}{k_b!(ns-k_b)!}, \tag{3.5}$$

for $b = 1 : nb$ bins.

SKETCH OF PROOF. An indirect, hand-waving argument is that there are nb bins into which to sort sample data of ns simulation events, such that any given event will end up in bin B_b with the probability, $p_b = f_b^{(\mathrm{th})}/ns$, and the observed frequency count for that bin, $f_b^{(\mathrm{sim})}$, will be a multinomial variate [**3, 8**].

The direct argument follows from the basic probability principles, multiple induction and the multinomial expansion. The IID property of the theoretical joint distribution yields a separable density for the simulation events,

$$\overline{\phi}^{(\mathrm{th})}(\boldsymbol{\eta}) = \prod_{s=1}^{ns}[\phi^{(\mathrm{th})}(\eta_s)] \ , \ \text{where} \ \int_{B_b} \phi^{(\mathrm{th})}(\eta_s) d\eta_s = p_b \ , \tag{3.6}$$

for any s, where $\boldsymbol{\eta} \equiv [\eta_i]_{ns\times 1}$, the realized η_s variable corresponding to the simulated random log-return $\Delta[\ln(S_s^{(\mathrm{sim})})]$ variable. The multinomial expansion can be written in the form:

$$\left(\sum_{b=1}^{nb} y_b\right)^{ns} = ns! \prod_{b=1}^{nb-1}\left[\sum_{f_b=1}^{ns-r_b} \frac{y_b^{f_b}}{f_b!}\right] \cdot \frac{y_{nb}^{ns-r_{nb}}}{(ns - f_{ns-r_{nb}})!} \ , \tag{3.7}$$

where $r_b = \sum_{d=1}^{b-1} f_d$ if $b > 1$, but $r_b \equiv 0$ if $b = 1$. The counting constraints are also operative: $\sum_{b=1}^{nb} k_b = ns = \sum_{b=1}^{nb} f_b = r_{nb+1}$.

Also, let $u_b(\boldsymbol{\eta}) \equiv \sum_{s=1}^{ns} U_b(\eta_s)$, where $U_b(\eta_s) \equiv U(\eta_s, B_b)$ is the indicator of b_b. Let the domain $R \equiv (-\infty, +\infty) = B_b + (R - B_b)$ be the real line with its partition for bin B_b and its complement $(R - B_b)$, whose segments have probabilities p_b and $(1 - p_b)$, respectively, with respect to the theoretical distribution. Since $U_b(\eta_s) = 1$ on bin B_b and $U_b(\eta_s) = 0$ on the bin complement $(R - B_b)$, a simpler, generalized form of $\phi^{(\mathrm{th})}(\eta_s)$ can be found, assuming F is a proper test function,

$$\begin{aligned} \int_R d\eta_s \phi^{(\mathrm{th})}(\eta_s) F(U_b(\eta_s)) &= \left(\int_{B_b} + \int_{R-B_b}\right) d\eta_s \phi^{(\mathrm{th})}(\eta_s) F(U_b(\eta_s)) \\ &= p_b F(1) + (1 - p_b) F(0) \ , \end{aligned}$$

So in the generalized sense,

$$\phi^{(\mathrm{th})}(\eta_s) \stackrel{\mathrm{gen}}{=} p_b \delta(U_b(\eta_s) - 1) + (1 - p_b)\delta(U_b(\eta_s)) \ ,$$

where $\delta(x)$ is the Dirac delta function, which will facilitate the application of a variation binomial theorem.

For $nb \geq 2$ bins, the derivation of the distribution of a single multinomial variate is the same,

$$\begin{aligned}
\Phi^{(\mathrm{sim})}(k_b) &\equiv \mathrm{Prob}\left[f_b^{(\mathrm{sim})} = k_b\right] = \mathrm{E}\left[\delta\left(k_b - f_b^{(\mathrm{sim})}\right)\right] \\
&= \prod_{s=1}^{ns}\left[\int_R d\eta_s \phi^{(\mathrm{th})}(\eta_s)\right]\delta\left(k_b - u_b(\boldsymbol{\eta})\right), \\
&= \prod_{s=1}^{ns}\left[\int_R d\eta_s\left(p_b\delta(U_b(\eta_s)-1) + (1-p_b)\delta(U_b(\eta_s))\right)\right]\delta\left(k_b - u_b(\boldsymbol{\eta})\right) \\
&= \sum_{s=1}^{ns}\left[\binom{ns}{s} p_b^s(1-p_b)^{ns-s}\prod_{r=1}^{s}\left[\int_R d\eta_r\delta(U_b(\eta_r)-1)\right]\right. \\
&\quad \left.\cdot\prod_{q=s+1}^{ns}\left[\int_R d\eta_q\delta(U_b(\eta_q))\right]\right]\delta(k_b - s) \\
&= \sum_{s=1}^{ns}\binom{ns}{s} p_b^s(1-p_b)^{ns-s}\delta(k_b - s) \\
&= ns!\frac{p_b^{k_b}(1-p_b)^{ns-k_b}}{k_b!(ns-k_b)!},
\end{aligned}$$

using the idea of the binomial expansion theorem in generalized form. Note that in the above equation, the discrete Kronecker delta $\delta_{k,s}$ has been interchanged with the continuous Dirac delta function $\delta(k-s)$. The derivation for the distribution for more than one free multinomial variate is similar, except in dimensional complexity.

It is more straight forward to derive the bin mean $\mu_b^{(\mathrm{sim})}$ and variance $(\sigma_b^{(\mathrm{sim})})^2$, but they are standard multinomial moments [**3, 8**]. □

REMARK 3.2. The results are independent of the underlying distribution, except for general properties, but in the sequel it is assumed that $f_b^{(\mathrm{sim})} = f_b^{(\mathrm{sp})}$ and $f_b^{(\mathrm{th})} = f_b^{(\mathrm{jd})}$.

3.3. Multinomial Maximum Likelihood Estimation. Let $\Phi^{(\mathrm{sim})}(\boldsymbol{k}) = \Phi^{(\mathrm{sp})}(\boldsymbol{k})$, $\Phi^{(\mathrm{th})}(\boldsymbol{k}) = \Phi^{(\mathrm{jd})}(\boldsymbol{k})$ and $f_b^{(\mathrm{th})} = f_b^{(\mathrm{jd})}(\boldsymbol{x})$, where $\boldsymbol{x}$ is the unknown parameter vector. The multinomial log-likelihood function used for the estimation has the form

$$\ln\left(\Phi^{(\mathrm{sp})}(\boldsymbol{k})\right) = \sum_{b=1}^{nb}\left[k_b\ln\left(f_b^{(\mathrm{jd})}(\boldsymbol{x})\right) - \ln(k_b!) - k_b\ln(ns)\right] + \ln(ns!), \tag{3.8}$$

subject to constraints $\sum_{b=1}^{nb} f_b^{(\mathrm{jd})}(\boldsymbol{x}) = ns = \sum_{b=1}^{nb} k_b$. However, given S&P500 input data, taking $k_b = f_b^{(\mathrm{sp})}$ as given, and ns is fixed, the only unknown is the parameter vector $\boldsymbol{x}$. The form of $f_b^{(\mathrm{jd})}(\boldsymbol{x})$ is known through the jump-diffusion density.

3.3.1. *Estimation Objective.* The essential part of the multinomial log-likelihood function, written as a minimization objective, is then

$$y(\boldsymbol{x}) \equiv -\sum_{b=1}^{nb}\left[f_b^{(\mathrm{sp})}\ln\left(f_b^{(\mathrm{jd})}(\boldsymbol{x})\right)\right], \tag{3.9}$$

with the change of sign to be suitable for minimization algorithms, neglecting constant parts, where $\boldsymbol{x}$ is the unknown jump-diffusion parameter vector and $\boldsymbol{f}^{(\mathrm{sp})}$ is the observed data vector.

Since $\mathrm{E}[f_b^{(\mathrm{sp})}] = f_b^{(\mathrm{jd})}(\boldsymbol{x})$, then

$$\mathrm{E}[y(\boldsymbol{x})] = -\sum_{b=1}^{nb}\left[f_b^{(\mathrm{jd})}(\boldsymbol{x})\ln\left(f_b^{(\mathrm{jd})}(\boldsymbol{x})\right)\right] .$$

So the mean objective is the entropy of jump-diffusion bin information.

3.4. Jump-Diffusion Moment Estimation Constraints. There are five (5) free jump-diffusion parameters:

$$\{\mu_{ld}, \sigma_d^2, \mu_j, \sigma_j^2, \lambda\} ,$$

for given Δt, as the reciprocal of the number of trading days per year. So, to reduce this set to a reasonable number, the multinomial maximum likelihood estimation is subjected to the mean and variance constraints:

3.4.1. *Mean and Variance Constraints.*

$$M_1^{(\mathrm{sp})} = M_1^{(\mathrm{jd})} \tag{3.10}$$

and

$$M_2^{(\mathrm{sp})} = M_2^{(\mathrm{jd})} , \tag{3.11}$$

where the theoretical jump-diffusion (jd) moments forms are given in (2.4-2.5) and the observed moments are given along with the definition of the S&P500 log-returns (3.1).

For the reduction to three (3) free parameters, the constraint eliminants are

$$\mu_{ld} = \left(M_1^{(\mathrm{sp})} - \lambda\Delta t\mu_j\right)/\Delta t \tag{3.12}$$

and

$$\sigma_d^2 = \left(M_2^{(\mathrm{sp})} - \lambda\Delta t\left(\sigma_j^2(1+\lambda\Delta t) + \mu_j^2\right)\right)/\Delta t , \tag{3.13}$$

the latter is subject to positivity constraints, for fixed and small $\Delta t \ll 1$.

3.5. Numerical Optimization: Golden Super Finder (GSF) and Nelder-Mead (NM). The numerical optimization procedures that will be tested here are the classical standard Nelder-Mead (MN) down-hill simplex method [**12**] and our own golden super finder (GSF) method [**4, 5, 6**]. The following is a point-by-point comparison of both methods:

- GSF is a multidimensional generalization of golden section search.
- GSF and NM are general methods: no derivatives are needed.
- GSF makes hypercube and other constraints implementable.
- GSF searches beyond initial domain subject to constraints since it uses boundary points too, unlike golden section search.
- GSF searches for uni-modal minimum within domain.
- GSF is more computational costly than Nelder-Mead's down-hill simplex direct search method.
- GSF is less likely to get caught in a local minimum so is more likely to find a global minimum than does NM, which is more sensitive to "lumpy" objectives like $y(x)$ with frequency data.

- The NM down-hill simplex method is available through popular programming systems such as MATLAB™ [**11**] under the `fminsearch` implementation and Numerical Recipes [**13**] under the `amoeba` implementation in Fortran, C and C^{++}.

4. Estimation Results and Discussion

Figure 1 is the histogram of empirical S&P500 closing log-returns $\Delta[\ln(S_s^{(\mathrm{sp})})]$ during the decade 1992-2001, using 100 bins. Note that most negative log-returns (crashes) have a larger magnitude than the positive ones (rallies), which typically leads to negative skew. Also, the tails starting from the shoulders of the most probable part of the distribution are much thicker than would be expected from a log-normal distribution, while the rare and large deviation lingering tails underscore the inadequacy of the log-normal distribution to model this aspect of the log-returns $\Delta[\ln(S_s^{(\mathrm{sp})})]$.

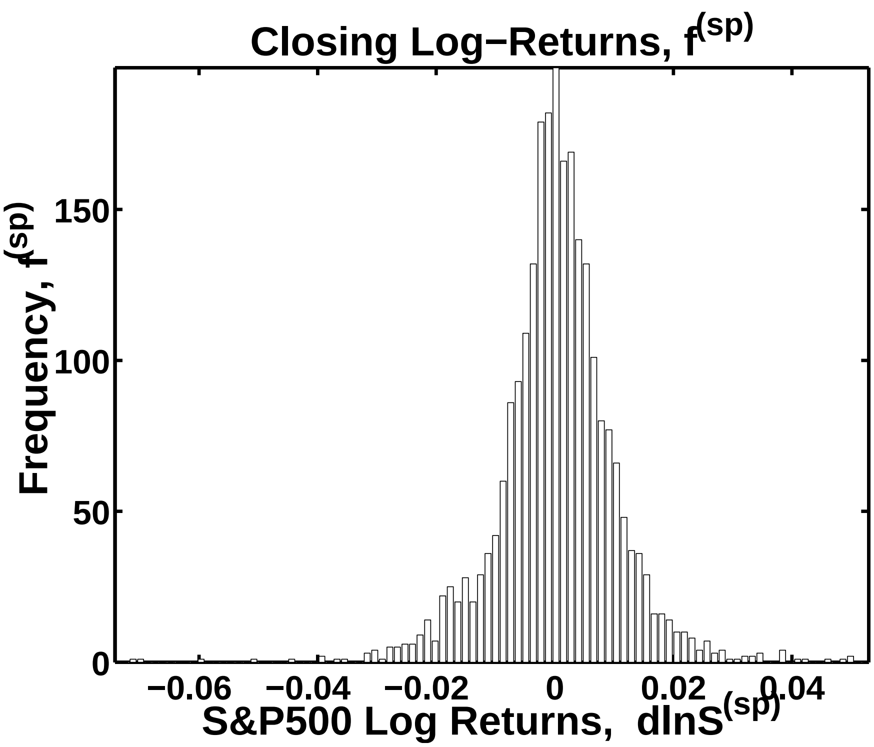
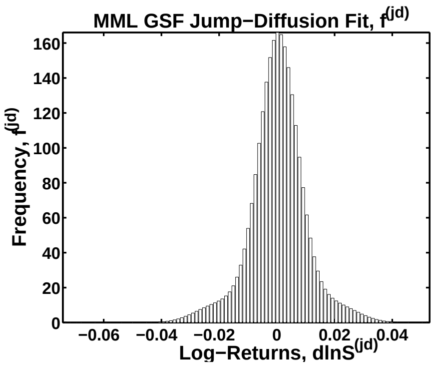
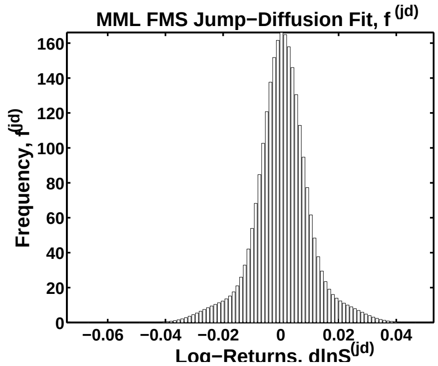


FIGURE 1. Histogram of Empirical S&P500 log-returns for the decade 1992-2001, using 100 bins.

In Figure 2, the fitted, predicted histograms are compared using our GSF method in the Subfig. 2(a) and the FMS method (MATLAB's `fminsearch` implementation of the Nelder-Mead algorithm) in Subfig. 2(b). This pair of figures should be viewed as a model of the underlying distribution for a moderate to large size simulation producing Fig. 1. More refined estimation methods would be needed to capture the extremes or outliers of the tails.



(a) Multinomial maximum likelihood estimation using GSF approximation.



(b) Multinomial maximum likelihood estimation using FMS approximation.

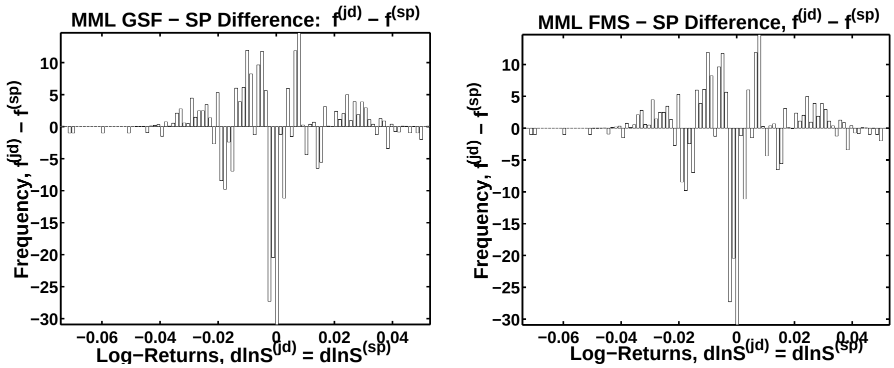
FIGURE 2. Comparison the histograms of predicted Log-Returns for the theoretical multinomial maximum likelihood (MML) estimation using (a) GSF optimization approximation and (b) Nelder-Mead (MATLAB `fminsearch` (FMS)) approximation. The parameter fittings for both are from the data of the empirical S&P500 histogram in Fig. 1 for the decade 1992-2001, using 100 bins.

In Figure 3, the theoretical to observed differences between the fitted, predicted histograms are compared using our GSF method Subfig. 3(a) and the FMS method (MATLAB's `fminsearch` implementation of the Nelder-Mead algorithm) Subfig. 3(b). These deviations of the theoretical distributions from the observed distribution in the pair of figures demonstrates that the GSF and FMS approximations to the multinomial maximum likelihood are very comparable.

In Figure 4, another comparison is given, this time comparing the multinomial maximum likelihood (MML) fit in Subfig. 4(a) and the weighted least squares (WLS) (Hanson and Westman [**7**]) method that corresponds to the usual maximum likelihood based on the normal distribution model in Subfig. 4(b). Both parameter fittings were computed using the same FMS method (MATLAB's `fminsearch` implementation of the Nelder-Mead algorithm). The WLS method places heavier emphasis on the tails. Note that there is a difference in scale between the left and right pair of figures.

In Figure 5, the theoretical to observed differences for the FMS-MML in Subfig. 5(a) and the FMS-WLS in Subfig. 5(b) approximate distribution estimates with respect to the S&P500 distribution. This pair of figures corresponds to the previous pair of figures in Fig. 4. Noting that the scales Subfig. 5(a) and Subfig. 5(b) pair of figures differ, but the FMS-WLS results in the (b) figure indicate that the WLS method significantly overestimates the near peak S&P500 values, especially on the positive side.

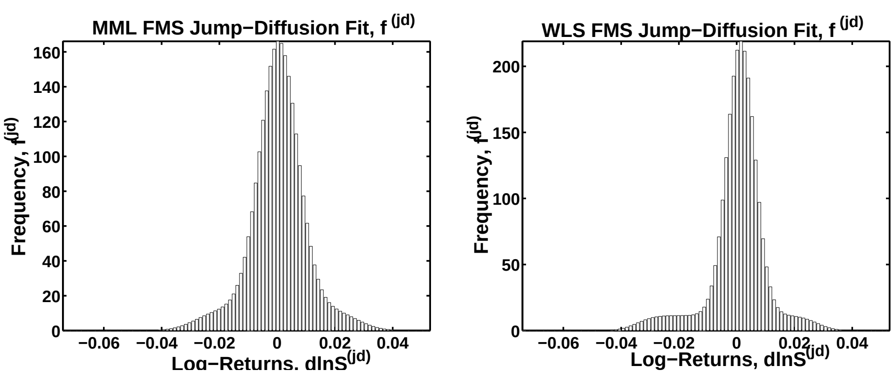
In these calculations for the histograms, although the theoretical log-uniform jump, log-normal diffusion distribution is on the fully infinite log-return domain $(-\infty, +\infty)$, the bins are necessarily of finite extent. However, in the display, the bin range includes all significant simulated probabilities in the main part and the tails



(a) Differences between MML GSF approximation and SP data.

(b) Differences between MML FMS approximation and SP data.

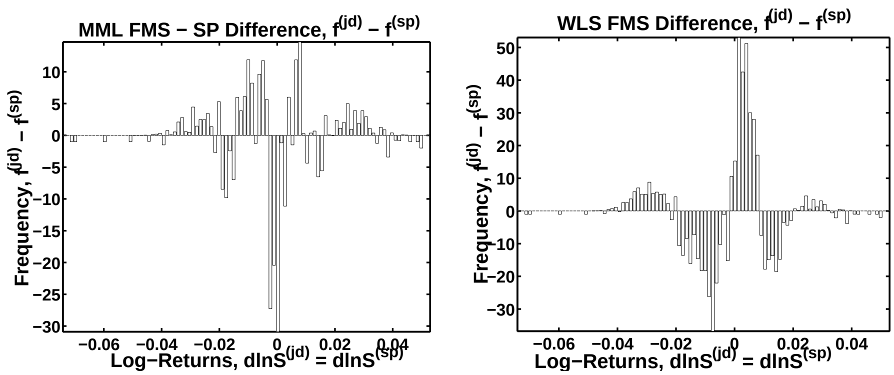
FIGURE 3. Comparison of the histogram differences for the theoretical multinomial maximum likelihood (MML) fit using (a) our GSF method and (b) FMS (`fminsearch`, the MATLAB implementation of the Nelder-Mead algorithm). Both differences are relative to the empirical S&P500 histogram in Fig. 1 for the decade 1992-2001, using 100 bins.



(a) Multinomial maximum likelihood estimation using FMS approximation.

(b) Weighted Least Squares estimation using FMS approximation.

FIGURE 4. Comparison of the histograms of predicted Log-Returns for the theoretical multinomial maximum likelihood (MML) fit using (a) Nelder-Mead (MATLAB `fminsearch` (FMS))and (b) the weighted least squares (WLS) [**7**] fit also using FMS. Both differences are relative to the empirical S&P500 histogram in Fig. 1 for the decade 1992-2001, using 100 bins.



(a) Differences between MML FMS approximation and SP data.

(b) Differences between WLS FMS approximation and SP data.

FIGURE 5. Comparison of the histogram differences for the multinomial maximum likelihood (MML) fit using (a) Nelder-Mead (MATLAB `fminsearch` (FMS)) and (b) the weighted least squares (WLS) [**7**] fit also using FMS. Both differences are relative to the empirical S&P500 histogram for the decade 1992-2001, using 100 bins.

of the distribution within double precision accuracy. In fact, the normal component of the distribution is already exponentially small where the uniformly distributed contributions from the secant-normal part of the distribution are still significant. The secant-normal part smooths and spreads out the log-uniform part of the distribution while fattening the tails where the pure normal part should be exponentially small. The negligibility of the tails was tested in the usual process of testing the code. In terms of exact representation the distribution may be thought to truncated, but in practical computational terms the distribution is fully represented within computer precision.

In Table 1, there is a comparison summary of derived distribution parameters for the log-normal-diffusion, log-uniform-jump distribution by the maximum likelihood method. The jump mean μ_j and standard deviation σ_j parameters are much smaller than the corresponding diffusion values, μ_d and σ_d, but the jump values, μ_j, σ_j and λ do not scale with the trading day time scale. The jump rates λ are quite large for a yearly rate, but that includes the whole range of the uniform distribution, the tails plus the central part of the distribution where the smaller jumps would be hidden by the normal part of the distribution. The WLS values of the mean diffusive drift μ_d are about double the MML values and are about half the diffusive standard deviation σ_d, while WLS is closer on the jump values. The MML estimations serve as the better estimate of the observed values, since these first and second moment parameters can not be directly separated from the observed data, although the theoretical and observed overall first and second moments have been constrained to match each other.

In Table 2, the skew and kurtosis coefficients are compared to S&P500 values for the same parameter fitting methods, the multinomial maximum likelihood

TABLE 1. Comparison summary of derived distribution parameters for the log-normal-diffusion, log-uniform-jump distribution by the maximum likelihood (MML) method using both GSF and FMS optimal search procedures and by the weighted least squares (WLS) using just the FMS search procedure.

Method	μ_d	σ_d	μ_j	σ_j	λ
MML GSF	+0.142	0.0861	-7.32e-4	0.0159	56.1
MML FMS	+0.143	0.0862	-7.40e-4	0.0160	56.0
WLS FMS	+0.364	0.0435	-4.76e-3	0.0179	55.3

(MML) method using both GSF and FMS optimal search methods and the weighted least squares (WLS) method using the FMS procedure only. These are the same parameter fitting methods and optimal search procedure combinations used for the jump-diffusion first and second moment parameters in Table 1. Since these are coefficients of overall moments, not their diffusion and jump components, they can be compared to the observed S&P500 data. The MML estimates of the coefficients of skewness (normalized third moment) differ from the observed by about -47%, while the WLS estimates differ by about -58%. The MML estimates of the kurtosis coefficient (normalized fourth moment) differ by about +0.78% from the observed values, which is excellent and the difference is insignificant, while the WLS kurtosis coefficient estimates differ by +49%. Theoretical estimates of the third and fourth moments are an extreme numerical test for comparison to the observed data, but the results, though not great for the skewness coefficient, demonstrate that the proper maximum likelihood method for binned data gives better results than the more general purpose method of weighted least squares, presumed to be better than the standard, unweighted least squares. Our weighted least squares method is a maximum likelihood method based upon the exponent of a normal distribution, as least square methods can be interpreted, but the weights used here were those from the multinomial distribution [**7**].

TABLE 2. Skew and kurtosis coefficients compared to S&P500 values for the same fitting methods and optimal search procedures of Table 1.

Method	$\beta_3^{(\mathrm{jd})}$	$\beta_3^{(\mathrm{sp})}$	$\beta_4^{(\mathrm{jd})}$	$\beta_4^{(\mathrm{sp})}$
MML GSF	-0.153	-0.291	7.86	7.80
MML FMS	-0.155	-0.291	7.86	7.80
WLS FMS	-0.121	-0.291	11.6	7.80

In Table 3 is a comparison summary of computational performance measures in terms of the number of iterations, the number of objective function evaluations and timings on two different computer processors. The parameter fitting methods with corresponding computational numerical optimal search method combinations are the same ones used in Tables 1-2. Clearly, the classical FMS direct search method is superior in performance over our GSF method for computing the MML estimates, based upon the number of iterations, the number of function evaluations and the timings. Note that the number of function evaluations correspond better to computational costs as indicated by the timing values, so they are better computational measures than the number of iterations. However, not shown here, our experience shows that our GSF can be more robust where the observed binned data is not too smooth, i.e., lumpy, so that FMS would get stuck on extraneous local extrema due to data irregularities.

TABLE 3. Comparison summary of computational performance measures:

Method Used	Number of Iterations	Function Evaluations	Timings (slow, secs.)	Timings (fast,secs.)
MML GSF	13	832	1.38e+4	3.70e+3
MML FMS	43	101	1.68e+3	3.74e+2
WLS FMS	75	146	2.73e+3	6.86e+2

Legend:

- MML GSF = Multinomial maximum likelihood (MML) using golden super finder.
- MML FMS = Multinomial maximum likelihood (MML) using Nelder-Mead (FMS).
- WLS FMS = Weighted least squares (WLS) using Nelder-Mead (FMS).
- Using same tolerances: tolx = 0.5e-4 and toly = 0.5e-3, using the same initial $\boldsymbol{x}$.
- Slow means P2@400MHz and fast means P4@2GHz CPUs or processors.

5. Conclusions

- It has been shown that the distribution of simulation bin frequencies is multinomial and hence leads to proper maximum likelihood estimation of the theoretical simulation distribution. The multinomial maximum likelihood is the proper one for binned data in general, and not just for the jump-diffusion model used here.
- Multinomial maximum likelihood estimates of jump-diffusion parameters are similar using either Nelder-Mead down-hill simplex or golden super finder, although Nelder-Mead is much faster.
- Multinomial maximum likelihood estimates of the skewness coefficient differ by -47% from the observed values, while the estimate of the kurtosis coefficient differ by a very small amount, +0.78%, giving strong support for the uniform distribution and the multinomial estimation procedure.
- Weighted least squares estimates for jump-diffusion parameters have greater sensitivity to jump rare events (outliers), but it does much worse on skew and kurtosis coefficients, exhibiting much greater differences.
- **Future Considerations:**
  - Study other jump amplitude distributions, with multinomial maximum likelihood estimation, that are more sensitive to the rare jump, outlier tails in real market distributions.
  - Introduce stochastic volatility into model (See Andersen, et al. [**1**] who found jumps alone were insufficient and substantiated the need for stochastic volatility, but they used an extraordinarily large parameter space and a not so appropriate jump-amplitude distribution).

Acknowledgement

The first two authors wish to thank George Yin and Qing Zhang for organizing such a splendid AMS-IMS-SIAM Joint Summer Research Conference on the Mathematics of Finance at the Snowbird Resort in Utah and providing conference support.

References

[1] T. G. Andersen, L. Benzoni, and J. Lund, *An empirical investigation of continuous-time equity return models,* J. Finance, **57**(3) (2002), 1239-1284.
[2] F. Black and M. Scholes, *The Pricing of Options and Corporate Liabilities*, J. Political Economy, **81** (1973), 637-659.
[3] M. Evans, N. Hastings and B. Peacock, *Statistical Distributions*, 3rd ed., John Wiley, New York, NY, 2000.
[4] F. B. Hanson and J. J. Westman, *Optimal consumption and portfolio control for jump-diffusion stock process with log-normal jumps*, Proc. 2002 Amer. Control Conf., 2002, pp. 4256–4261
[5] F. B. Hanson and J. J. Westman, *Portfolio optimization with jump–diffusions: Estimation of time-dependent parameters and application*, Proc. 2002 Conf. Decision and Control, 2002, pp. 377-382.
[6] F. B. Hanson and J. J. Westman, *Jump-diffusion stock return models in finance: Stochastic process density with uniform-jump amplitude*, Proc. 15th Int. Sympos. Mathematical Theory of Networks and Systems, 2002, pp. 1–7.
[7] F. B. Hanson and J. J. Westman, *Jump-diffusion stock-return model with weighted fitting of time-dependent parameters*, Proc. 2003 Amer. Control Conf., 2003, pp. 4869-4874.
[8] F. B. Hanson and J. J. Westman, *Applied Stochastic Processes and Control for Jump-Diffusions: Modeling, Analysis and Computation* SIAM Books, Philadelphia, PA, to appear 2004.
[9] S. G. Kou, *A jump diffusion model for option pricing*, Management Science, **48** (2002), 1086-1101.
[10] R. C. Merton, *Continuous-Time Finance*, Basil Blackwell, Cambridge, MA, 1990.
[11] C. Moler et al., *Using MATLAB, Version 6*, Mathworks, Natick, MA, 2000.
[12] J. A. Nelder and R. Mead, *A simplex method for function minimization*, Computer Journal, **7** (1965), 308–313.
[13] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in C++: the Art of Scientific Computing*, Cambridge U. Press, UK, 2002.
[14] H. M. Taylor and S. Karlin, *An Introduction to Stochastic Modeling*, 3rd ed., Academic Press, New York, NY, 1998.
[15] Yahoo! Finance, *Historical Quotes, S & P 500 Symbol^SPC*, URL:http://chart.yahoo.com/. February 2002.

Department of Mathematics, Statistics, and Computer Science, M/C 249, University of Illinois, Chicago, IL 60607-7045
E-mail address: hanson@math.uic.edu

Department of Mathematics and Statistics, Miami University, Oxford, Ohio 45056
E-mail address: westmajj@muohio.edu

Department of Mathematics, Statistics, and Computer Science, M/C 249, University of Illinois, Chicago, IL 60607-7045
E-mail address: zzhu@math.uic.edu