Estimating Jump Diffusion Structural Credit Risk Models

Hoi Ying Wong∗
Department of Statistics,
The Chinese University of Hong Kong,
Shatin, N.T., Hong Kong

Chin Pang Li
Department of Risk Management,
Fubon Bank, Hong Kong

August 21, 2006

∗Corresponding author: fax: (852) 2603-5188; e-mail: hywong@cuhk.edu.hk.
Estimating Jump Diffusion Structural Credit Risk Models

Abstract

There is strong evidence that structural models of credit risk significantly underestimate both credit yield spreads and the probability of default if the value of corporate assets follows a diffusion process. Adding a jump component to the firm value process is a potential remedy for the underestimation. However, there are very few empirical studies of jump-diffusion (or Levy) structural models in the literature. The major challenge is the estimation of hidden variables, such as the firm value, volatility, and parameters of the jump component, as the value of corporate assets is not directly observable. In practice, parameters and the value of the firm should be estimated using the market values of equities. This paper provides a promising estimation method for jump-diffusion processes in structural models that are based on observed stock data. We show that the traditional estimation methods for structural models, the variance-restriction method and maximum likelihood estimation, fail when jumps appear in credit risk models. We then propose a penalized likelihood approach and devise a corresponding expectation-maximum algorithm. The approach is applied to the jump-diffusion processes of Merton (1976) and Kou (2002) and the performance is examined through a series of simulations and empirical data.

JEL classification: To be determined

Keywords: Credit Risk, Jump-Diffusion, Quasi-Bayesian, Maximum Likelihood
1 Introduction

The measurement of credit risk is of natural interest to financial practitioners, regulators, and academics. An accurate credit risk model is essential for sound risk management, for evaluation of the vulnerability of lender institutions, and for pricing credit derivatives. In the banking industry, the regulatory framework [Basel (2004)] encourages the active involvement of banks in assessing the likelihood of defaults. The need for an accurate and practical credit risk framework has grown exponentially in recent years. A desirable approach should be consistent with both economic theory and empirical investigations.

The finance literature has produced a variety of models to measure default risk. Most of these are structural models based on the economic theory of capital structures and the incentive issues among equity holders, debt holders, management, and other stakeholders in the corporation. More precisely, structural models focus on the stochastic process of the value of corporate assets and postulate that default occurs when the firm value hits a threshold value.

Structural models can be divided into a barrier-independent group and a barrier-dependent group. The former group considers that default occurs at some specified discrete time points. For instance, Merton (1974) assumes that default only occurs at debt maturity and views the market value of equity as a standard call option on the firm’s assets. Geske (1977) uses a compound option approach to describe default at multiple time points. For the barrier-dependent group, default is allowed at any time before the debt maturity. This idea originates from Black and Cox (1977) who introduce the concept of default barrier (distress level). Whenever the downside barrier is hit, debt holders will exercise their right to pull the plug and force equity holders to declare bankruptcy before the firm value deteriorates further. Extensions on barrier-dependent models include Longstaff and Schwartz (1995, LS), Lelant and Toft (1996, LT), Collin-Dufresne and Goldstein (2001, CDG), and others. Apart from academic research, structural models have been put into commercial use, firstly by Moody’s KMV.

There is strong empirical evidence, however, that structural models significantly underestimate credit yield spreads and the probability of default. Jones, Mason and Rosenfeld (1984) find that the predicted bond yield is too low in Merton’s (1974) model. The problem is more severe for non-investment grade bonds. Ogden (1987) finds a similar result using newly issued bonds. Eom et al. (2004) empirically test the Merton, Geske, LS, LT, and CDG models and find that these models generate a very large predictive error in terms of credit spread. Tarashev (2005) observes that structural models produce a probability of default that is sig-
nificantly less than the empirical default rate.

Although the performance of structural models can be improved by utilizing a better implementation method, the problem of underestimation still exists. Duan (1994) points out that the traditional variance-restriction (VR) method for the Merton model is inconsistent with the model feature. He then proposes maximum likelihood (ML) estimation as an alternative implementation method. Duan and Simonato (2002) document that the ML estimation in conjunction with the Merton model improve the estimate of deposit insurance values over the traditional approach. Ericsson and Reneby (2005) use a series of simulations to further ensure that ML estimation outperforms the VR method for both barrier-independent and barrier-dependent models. Li and Wong (2006) find empirical evidence that the ML estimation effectively improves the performance of the Merton, LS and CDG models. However, these models still suffer from significant underestimation in credit spread for short-term bonds and low rating bonds.

The underestimation of credit risk with structural models was first recognized by Merton (1974). The key reason is that a sudden drop in the firm value is impossible under the diffusion process; firms never default by surprise. Therefore, a jump-diffusion process is a remedy for the structural models. Zhou (2001) reveals that incorporating jumps into the asset value process can generate realistic shapes for the term structure of credit spread, such as upward sloping, flat, hump-shape, and downward sloping. In diffusion models, some of the shapes are impossible. Hilberink and Rogers (2002) extend the LT model using Levy processes, which only allow for downward jumps in the firm value. Chen and Panjer (2003) connect jump-diffusion structural models with reduced-form models. Dao and Jeanblanc (2006) consider the double exponential jump-diffusion in a barrier-dependent model.

Another reason for jump-diffusion structural credit risk model is that empirical evidence on jumps in financial assets is abundant (see, e.g., Bates, 1996; Andersen et al., 2002; Pan, 2002; Eraker et al., 2003). In structural models, when the equity price exhibits a jump, the firm value is also expected to jump. Using a regression-based analysis, Zhang et al. (2005) observe that credit default swaps are sensitive to jumps on equity returns. Delianedis and Geske (2001) find some evidence that jump risk is a component of the corporate credit spread. Huang and Huang (2003) use a double exponential jump-diffusion model and find that it accounts for a small part of the credit risk. Unfortunately, none of these works estimate

---

1This widely adopted approach has been to solve a system of equations that match the observed stock prices and estimated stock volatility with model outputs.
the jump-diffusion process directly. Instead, the firm value and volatility are first obtained from the VR approach, and the jump distribution is then fitted to the residual credit spread. This ad hoc approach ignores the interaction among the firm value, volatility, and the jump component in the estimation procedure. In fact, to our knowledge, no empirical work has been devoted to examining the jump component in the firm value process directly.

The fundamental difficulty in carrying out an empirical study of a jump-diffusion structural model is the estimation of hidden variables, such as the firm asset value, the drift and volatility of the diffusion component, and the parameters of the jump component, from equity returns, because the value of corporate assets is not directly observable. This motivates us to propose in this paper an efficient estimation method for jump-diffusion structural credit risk models. We focus on the log-normal jump-diffusion process of Merton (1976, MJD) and the double exponential jump-diffusion process of Kou (2002, KJD) due to their popularity and analytical tractability. However, the proposed estimation is not limited to these two cases.

We propose the penalized likelihood estimation (PLE) for jump-diffusion structural models and stress that it is not an obvious extension of its diffusion counterpart. In fact, both the VR approach and the ML estimation fail to estimate jump-diffusion structural models. Specifically, the VR approach which solves the firm value and volatility from two simultaneous equations cannot accomplish the estimation of additional parameters; whereas, the likelihood function for equity data degenerates over the parameter space of jump-diffusion processes. In other words, the likelihood function can reach infinity for some sets of observed equity data. In such a situation, the maximum likelihood estimator (MLE) cannot be defined. We also stress that the proposed PLE is not an obvious extension of jump-diffusion processes for observable data. As the value of corporate assets is not directly observable, estimation should be based on the high frequency data of equity returns. When the firm asset value evolves as a jump-diffusion process with constant parameters, the process for the equity return can be very complicated.

Business and risk management applications of PLE can be found in Hamilton (1991) and Venkataraman (1997), respectively. Moreover, PLE has been widely used in estimating mixture distributions and, in many cases, proven to be efficient and stable (see, e.g., Yu et al, 1994; Mammen and van de Geer, 1997; Eggermont and LaRicca, 2001). To our knowledge, this is the first paper to bring PLE to the corporate credit risk literature. We are also the first to propose an estimation method for the jump-diffusion structural models of credit risk.

The intention of the proposed approach is to penalize the likelihood function
for equity returns with prior distributions such that the penalized function never blows up to infinity at all points of the parameter space. Parameters are then obtained by maximizing the penalized likelihood, subject to the constraint that observed equity prices should match the model outputs. We construct PLE for structural models under both the MJD and KJD processes. In a nutshell, the proposed estimation combines the ideas of classical penalized likelihood and the ML estimation of Duan (1994). To aid implementation, we establish an expectation-maximum (EM) algorithm for obtaining estimates from the PLE approach. EM algorithm is not a new concept to the financial market. As pointed out by Duan et al. (2004), the KMV methodology is equivalent to the ML estimation of Duan (1994) in the sense that the former is an EM algorithm of the latter. Our simulation shows that the EM algorithm for the proposed PLE is efficient and accurate.

The remainder of the paper is organized as follows. Section 2 reviews several jump-diffusion structural credit risk models. Section 3 introduces the proposed estimation method. Here, we also prove the degeneracy of the likelihood function for equity returns and establish the penalized likelihood estimation. Section 4 devises EM algorithms for the proposed approach. Section 5 shows the performance of the estimation with Monte Carlo simulations and empirical data. Section 6 concludes the paper. Some technical details are presented in the Appendix.

2 Jump-Diffusion Structural Credit Risk Models

The value of corporate assets, \( V \), is assumed to follow a jump-diffusion process under the physical probability measure, \( \mathbb{P} \):

\[
\frac{dV(t)}{V(t^-)} = \mu dt + \sigma dW_t + d \left( \sum_{j=1}^{N(t)} (Z_j - 1) \right),
\]

where \( \mu \) is the drift of the business, \( \sigma \) the asset value volatility, \( W_t \) a Wiener process, \( N(t) \) a Poisson process with intensity \( \lambda \), and \( \{Z_j > 0\} \) a sequence of independent identically distributed (i.i.d.) random variables. The distribution of \( Y_j = \log Z_j \) will be specified later. We postulate that \( W_t, N(t) \) and \( Y_j \) are independent and all of the parameters are real constants. Applying Itô’s lemma to \( \omega(t) = \log V(t) \) yields:

\[
d\omega(t) = \left( \mu - \sigma^2/2 \right) dt + \sigma dW_t + Y dN(t), \quad \omega(0) = \log V(0).
\]
Although the proposed estimation method is applicable to a general jump-diffusion process, we pay special attention to the MJD and KJD processes for the purpose of illustration. The MJD process assumes $Y$ to have a normal distribution with the density function:

$$f_Y(y) = \frac{1}{\sqrt{2\pi s^2}} \exp \left[ -\frac{(y - k)^2}{2s^2} \right],$$

whereas the KJD process assumes $Y$ to have a double exponential distribution with the density function,

$$f_Y(y) = p\eta_1 e^{-\eta_1 y} 1_{\{y \geq 0\}} + (1 - p)\eta_2 e^{\eta_2 y} 1_{\{y < 0\}}, \ p \geq 0, \ \eta_1 > 1, \ \eta_2 > 0. \ (4)$$

In credit risk management, the fundamental interest is the measurement of the probability of default (PD) of a credit obligator. Financial economists used to model default as the event in which the firm value drops below a distress level. Suppose the firm value evolves as a jump-diffusion process, as in (1). Then the key to measuring PD is the estimation of the firm’s asset-value and the parameters of the process. For the purpose of estimation, we need the probability density function of $\omega_i$ conditioning on $\omega_{i-1}$, where $\omega_i = \omega(t_i)$ represents the log-asset value at time $t_i$. When the time difference $\Delta t_i = t_i - t_{i-1}$ is small, by the properties of the Poisson process $N(t)$, we know that $P(\Delta N_i = 0) = 1 - \lambda \Delta t_i + o(\Delta t_i)$, $P(\Delta N_i = 1) = \lambda \Delta t_i + o(\Delta t_i)$ and $P(\Delta N_i > 1) = o(\Delta t_i)$, where $\Delta N_i = N(t_i) - N(t_{i-1})$. In other words, there is at most one jump arriving in a short period of time, because the probability of additional jumps is negligible. If there are no jumps in the period, then $\omega_i|\omega_{i-1}$ follows a normal distribution; otherwise, it has a distribution derived from the convolution between the normal distribution and the distribution of the jump size. Thus, $\omega_i|\omega_{i-1}$ can be thought of as a random variable that has a mixture distribution for a small $\Delta t_i$ and the density function:

$$g(\omega_i|\omega_{i-1}) = (1 - \lambda \Delta t_i) f_X(\omega_i|\omega_{i-1}) + \lambda \Delta t_i f_{X+Y}(\omega_i|\omega_{i-1}), \ (5)$$

where $X$ and $Y$ represent the diffusion component and the jump size, respectively, so that

$$f_X(x|\omega_{i-1}) = \frac{1}{\sqrt{2\pi \sigma^2 \Delta t_i}} \exp \left[ -\frac{(x - \omega_{i-1} - \tilde{\mu})^2}{2\sigma^2 \Delta t_i} \right],$$

$$f_{X+Y}(\xi|\omega_{i-1}) = \int_{-\infty}^{\infty} f_X(\xi - y|\omega_{i-1}) f_Y(y) dy,$$

$$\tilde{\mu} = (\mu - \sigma^2/2) \Delta t_i.$$
In the MJD process, \((X + Y)|\omega_i - 1 \sim \mathcal{N}(\omega_i - 1 + \hat{\mu} + k, \sigma^2 \Delta t_i + s^2)\), and hence \(f_{X+Y}(\xi|\omega_i - 1)\) is a normal density. For the KJD process, a simple calculation yields,

\[
f_{X+Y}(\xi|\omega_i - 1) = p \cdot \eta_1 e^{-\frac{\eta_1(\xi - m_1)}{\Delta t_i}} \Phi\left(\frac{\xi - m_1 - \eta_1 \sigma^2 \Delta t_i / 2}{\sigma \sqrt{\Delta t_i}}\right)
+ (1 - p) \cdot \eta_2 e^{\eta_2(\xi - m_2)} \Phi\left(\frac{m_2 - \xi - \eta_2 \sigma^2 \Delta t_i / 2}{\sigma \sqrt{\Delta t_i}}\right),
\]

where

\[
m_1 = \omega_i - 1 + \hat{\mu} + \eta_1 \sigma^2 \Delta t_i / 2,
\]

\[
m_2 = \omega_i - 1 + \hat{\mu} - \eta_2 \sigma^2 \Delta t_i / 2.
\]

Let \(\theta\) be the vector of parameters of the process (1), i.e., \(\theta = (\mu, \sigma^2, \lambda, k, s^2)\) for the MJD process and \(\theta = (\mu, \sigma^2, \lambda, p, \eta_1, \eta_2)\) for the KJD process. The density function (5) leads to the likelihood function for the log-asset values:

\[
L^V(\theta) = \prod_{i=1}^{n} g(\omega_i|\omega_i - 1, \theta).
\]

Often, statistical inference can be drawn from the likelihood function and the observed data. Unfortunately, log-asset values are not directly observable in the market. In this paper, the firm value and the parameter \(\theta\) are filtered from stock price data using a structural framework. For the moment, we present the relationship between the firm value and stock price in view of the structural approach.

### 2.1 Barrier-Independent Model

Consider a simple capital structure of a firm that includes both equity and debt. Merton (1974) argues that equity holders are residual claimants on the firm’s assets after all obligations have been met. Upon the debt maturity, equity holders will receive the firm’s asset value less the debt if the firm is able to pay back loans. Otherwise, they will get nothing and the firm will declare bankruptcy. Therefore, the equity of a firm, \(S\), can be viewed as a standard call option on the firm’s assets with the strike price set to the book value of corporate liabilities, \(K\).

It is clear that the call option can be valued by discounting the expected payoff under the risk-neutral measure, \(Q\),

\[
S = e^{-rT} E^Q [ (V(T) - K)^+ | V ] ,
\]
where a constant interest rate $r$ is used for the purpose of illustration. Merton (1974) employs the Black-Sholes (1973) option pricing formula to establish an explicit relationship between the stock price and the firm value.

In our case, we need the call option pricing formula under the jump-diffusion process of (1). For both MJD and KJD processes, closed form solutions are available and summarized as follows:

1. For the MJD process, Merton (1976) has derived the closed form solution,

$$S = C_M(V; K, T, \theta^*) = \sum_{n=0}^{\infty} \frac{e^{-\lambda' T} (\lambda')^n}{n!} \cdot C_{BS}(V, K, \sigma^2_n, r_n, T), \quad (9)$$

where $\theta^* = (\mu^*, \sigma^2, \lambda^*, k^*, (s^*)^2)$ is a vector of risk-neutral parameters, $C_{BS}$ is the Black-Scholes pricing formula and

$$\lambda' = \lambda^*(1 + k^*), \quad r_n = r - \lambda^* k^* + \frac{n}{T} \log(1 + k), \quad \sigma^2_n = \sigma^2 + n(s^*)^2/T.$$ 

2. For the KJD process, Kou (2002) has derived the closed form solution for the call option as an infinite series of Hh functions, and Kou and Wang (2004) have obtained the Laplace transform of the option price. We only present the latter because of its computational efficiency.

$$S = C_K(V; K, T, \theta^*) = \mathcal{L}_{\tau, \zeta}^{-1} \left[ e^{-rT} \frac{V^{\zeta+1}}{\zeta (\zeta + 1)} e^{G(\zeta+1)T} \right]. \quad (10)$$

where $\theta^* = (\mu^*, \sigma^2, \lambda^*, p^*, \eta_1^*, \eta_2^*)$ is a risk-neutral parameter vector, $\mathcal{L}_{\tau, \zeta}^{-1}$ is the Laplace inversion with respect to $\zeta$ evaluated at $\tau = \log K$, and

$$G(x) = x \left( \mu^* - \frac{\sigma^2}{2} \right) + \frac{\sigma^2}{2} x^2 + \lambda^* \left( \frac{p^* \eta_1^*}{\eta_1^* - x} + \frac{(1 - p^*) \eta_2^*}{\eta_2^* + x} - 1 \right). \quad (11)$$

Under the jump-diffusion process of (1), Merton (1976) has pointed out that perfect hedging is impossible, but the delta-hedging strategy of Black and Scholes (1973) can hedge away the diffusion risk. The hedging strategy gives us a risk-neutral measure under which the parameters are the same as the physical counterparts, except that $\mu^* = r - \lambda k$. In fact, there are many possible risk-neutral
measures. Using the HARA utility for a representative agent, Kou (2002) clarifies the relationship between the physical parameter $\theta$ and the risk-neutral parameter $\theta^*$. Specifically, $\lambda^* = \lambda E(Z^{\alpha-1}), \mu^* = r - \lambda^* E(Z^\beta)$ and $Y^* = \beta Y_j$, where $\alpha \in [0, 1]$ is the power parameter of the HARA utility function, $\beta \in (-\infty, \infty)$ is an arbitrary constant and $Y^*$ is the log jump size under the risk-neutral measure. Hence, the risk-neutral process recognized by Merton is a special case of Kou’s in which $\alpha = \beta = 1$. Therefore, in addition to the process for the firm asset value, a complete jump-diffusion structural model should specify the utility function and the parameter $\beta$. To simplify matters, we adopt the risk-neutral measure of Merton (1976), i.e., $\alpha = \beta = 1$, in this paper. It means that we use the risk-neutral utility function ($\alpha = 1$) for all investors and that they only hedge the diffusion risk using the option delta. However, our estimation scheme is not limited to this consideration.

2.2 Barrier-Dependent Model

Barrier-dependent models are introduced by Black and Cox (1976). Their idea is to include a barrier, $H$, to trigger default so that the market value of equity is viewed as a down-and-out call (DOC) option on the value of corporate assets. Specifically,

$$S = \text{DOC}(V; T, K, H, \theta^*) = e^{-rT} E^Q \left[ (V(T) - K)^+ 1_{\{\min_{0 \leq t \leq T} V(t) > H\}} \right].$$

Longstaff and Schwartz (1995) extends the Black-Cox model to value corporate coupon bond under stochastic interest rates. However, they do not mention the modeling of equity. Therefore, it is reasonable to consider Black and Cox’s original proposal in modeling the stock price.

The barrier provision in options admits no closed form solution for most jump-diffusion processes, including the MJD process. The valuation usually relies on numerical methods. Metwally and Atiya (2002) propose an efficient simulation algorithm for barrier options under jump-diffusion processes. Their idea is to generate jump instants in advance. In between two successive jump instants, the asset value evolves as a Brownian motion so that the asset value just before the next jump instant can be generated using the Brownian bridge with the condition that the asset value does not hit the barrier in the time interval. This method is applicable to both MJD and KJD processes.

The KJD process has analytical tractability in pricing path dependent options. Kou and Wang (2004) derive a closed form solution for barrier options in terms
of Laplace transform. Kou et al. (2005) develop an analytical solution for the up-and-in call option in double Laplace transform. We modify the latter result to price the DOC option.

\[
\text{DOC}_K(V; T, K, H, \theta^*) = \mathcal{L}^{-1}_{T, \gamma} \mathcal{L}^{-1}_{r, \zeta} \left[ \frac{H^{\zeta+1} \eta_2^2 A(r+\gamma) + B(r+\gamma)}{\zeta (\zeta + 1)} \right],
\]

where \( \tau = \log K \), \( G(x) \) is defined in (11),

\[
A(r + \gamma) = \frac{(\eta_2^2 - \beta_{3,r+\gamma})(\beta_{4,r+\gamma} - \eta_2^2)}{\eta_2^2 (\beta_{4,r+\gamma} - \beta_{3,r+\gamma})} \left[ \left( \frac{H}{V} \right)^{\beta_{3,r+\gamma}} - \left( \frac{H}{V} \right)^{\beta_{4,r+\gamma}} \right],
\]

\[
B(r + \gamma) = \frac{(\eta_2^2 - \beta_{3,r+\gamma})}{(\beta_{4,r+\gamma} - \beta_{3,r+\gamma})} \left( \frac{H}{V} \right)^{\beta_{3,r+\gamma}} + \frac{(\beta_{4,r+\gamma} - \eta_2^2)}{(\beta_{4,r+\gamma} - \beta_{3,r+\gamma})} \left( \frac{H}{V} \right)^{\beta_{4,r+\gamma}},
\]

and \(-\beta_{3,r+\gamma}\) and \(-\beta_{4,r+\gamma}\) are the two negative roots of the equation,

\[
G(x) = r + \gamma,
\]

where \( G(x) \) is defined in (11). Explicit expressions of the roots appear in Appendix B of Kou et al. (2005).

We shall see that the quality and computational efficiency of the proposed estimation method can be greatly enhanced by an analytical solution for the equity price. Although it is still feasible to implement our approach with an efficient simulation method, the estimates would be subject to the discretization bias, and the execution time would be much longer. This is why we present the details on option pricing with the jump-diffusion processes.

3 Proposed Estimation

After specifying the default structure (whether or not it is barrier-dependent) and the process of the firm asset value, there is a one-to-one relationship between the firm asset value and the stock price. The previous section demonstrated some possible specifications. We use the notation, \( S(t) = h(t, V, K, T; \theta^*) \), to indicate the relationship. As the stock price must be an increasing function of \( V \), we must have a positive delta, i.e., \( \partial h / \partial V > 0 \). Given this fact and the process of \( V \), we are able to express the likelihood function of stock prices, \( L^S \), in terms of the likelihood function of the firm values by utilizing the arguments of Duan (1994).
Let $S = \{S_j = S(t_j) : j = 0, 1, \cdots, N\}$ be a set of $N + 1$ observed stock prices and $f(\log S_j | \log S_{j-1}, \theta)$ be the density function of $\log S_j$ conditioning on $\log S_{j-1}$. Then, the likelihood function for the log stock price is given by:

$$L^S(\theta) = \prod_{j=1}^{N} f(\log S_j | \log S_{j-1}, \theta).$$

As the density function of the firm’s asset values is known, we can express the density function for $\log S$ in the following way,

$$f(\log S_j | \log S_{j-1}, \theta) = g(\log V_j | \log V_{i-1}, \theta) \left| \frac{\partial \log S}{\partial \log V} \right|_{t=t_i}^{-1} \frac{\partial \log S}{\partial \log V} \left| \frac{\partial \log V}{\partial V} \right|_{t=t_i}^{-1}.$$

Hence, the likelihood function for equity data is given by,

$$L^S(\theta) = L^V(\theta) \prod_{i=1}^{N} \frac{S_i}{V_i} \left[ \frac{\partial h}{\partial V} \right]^{-1}.$$

and the log-likelihood function is

$$\ell^S(\theta) = \log L^V(\theta) + \sum_{i=1}^{N} (\log S_i - \omega_i) - \sum_{i=1}^{N} \log \left[ \frac{\partial h}{\partial V} \right] \left| \frac{\partial h}{\partial V} \right|_{t=t_i}^{-1}.$$

Under the diffusion process, Duan (1994) suggests obtaining $\theta$ by maximizing the log-likelihood function (13), subject to the constraint that

$$S_j = h(t_j, V_j, K, \theta^*), j = 0, 1, \cdots, N.$$

This method works very well under diffusion processes (see comments by Duan and Simonato, 2002; Ericsson and Reneby, 2005; Li and Wong, 2006). For jump-diffusion structural models, we unfortunately find that the log-likelihood function (13) diverges to infinity for some observed stock prices, and hence fails to estimate parameters.
3.1 Degeneracy of the Likelihood Function

There are two problems with the ML estimation when the jump component is incorporated into the firm value process. The first is that the likelihood function may diverge to infinity for some observed stock data. The second is that, when the likelihood diverges to infinity, the value of the firm asset volatility, $\sigma$, tends to zero which violates the model assumption. We formally write it as a proposition.

Consider $\theta = (\theta_X, \theta_Y)$, where $\theta_X = (\mu, \sigma^2, \lambda)$ collects the diffusion parameters and the Poisson process intensity, and $\theta_Y$ collects parameters of the distribution of $Y$. Let $\Theta_X$ be the parameter space of $\theta_X$, i.e. $\Theta_X = \{\theta_X : \sigma, \lambda > 0, \mu \in \mathbb{R}\}$, $\Theta_Y$ be the parameter space of $\theta_Y$, i.e. $\Theta_Y = \{\theta_Y : s \geq 0, k \in \mathbb{R}\}$ for the MJD process and $\Theta_Y = \{\theta_Y : \eta_1 > 1, \eta_2 > 0, p \in [0, 1]\}$ for the KJD process, $\Theta = \Theta_X \times \Theta_Y$ be the parameter space of $\theta$ and $\overline{\Theta}_X$ be the closure of $\Theta_X$. We introduce a set associated to the singularity $\sigma = 0$ given $N + 1$ observed stock prices. More precisely,

$$S_0(S) = \{\theta \in \overline{\Theta}_X \times \Theta_Y | \exists n \in \{1, ..., N\}, \tilde{\mu} = \log[h^{-1}(S_n)/h^{-1}(S_{n-1})], \sigma = 0\}.$$

**Proposition 3.1.** For any given data set $S$, the log-likelihood function, $\ell^S(\theta)$, defined by (13) degenerates at every point of $S_0$, i.e.,

$$\forall S \in (\mathbb{R}^+)^N, \exists S_0(S), \exists(\theta_m \in \Theta, m = 1, 2, \ldots), \lim_{m \to \infty} \theta_m = \theta', \lim_{m \to \infty} \ell^S(\theta_m) = \infty.$$  

**Proof:** See Appendix A.

As a consequence of Proposition 3.1, the maximum likelihood estimator cannot be defined for jump-diffusion processes. Moreover, the points that belong to $S_0(S)$ provide meaningless estimates for $\theta$. In particular, ML estimation fails to estimate either the MJD or the KJD structural model because it fails for all jump-diffusion structural models. In practice, when the ML estimation is used for jump-diffusion structural models, two problems may be encountered. Firstly, the optimization algorithm keeps running for a very long time without converging to a stable solution. This is because the global maximum is indeed infinity. Secondly, an unreasonably small volatility is obtained, because the volatility nears zero when the program is run for a long time.

3.2 Penalized Likelihood

The degeneracy of likelihood functions for mixture distributions is well known in the statistics literature, although the same cannot be said for jump-diffusion
structural models. A promising solution is the use of Bayesian or quasi-Bayesian techniques. In this paper, we propose a penalized likelihood estimation (PLE) that is in the quasi-Bayesian family. EM algorithms are then established to search the local maxima of the penalized likelihood function for different models. Note that the Bayesian approach which is usually implemented with Markov Chain Monte Carlo methods may require much a longer time period to get stable estimates. Moreover, Eggermont and LaRiccia (2001) show that the penalized likelihood estimator is strongly consistent and asymptotically efficient.

The proposed estimation consists of penalizing the likelihood function with prior distributions such that the penalized likelihood function never blows up to infinity. The prior distributions are essentially used to associate randomness to the parameters which cause the degeneracy of the likelihood function. Thus, it is important for us to identify all of the singularity points of the likelihood function. We find that all of the singularity points belong to the set $S_0(S)$. Equivalently, the log-likelihood function, $\ell^S(\theta)$, does not degenerate outside any neighborhood of $S_0(S)$. We denote the neighborhood as,$$S_\epsilon(S) = \{\theta \in \overline{\Theta} | \forall \theta' \in S_0(S), 0 = \sigma' \leq \sigma \leq \epsilon\},$$where $\Theta$ is the parameter space and $\overline{\Theta}$ is its closure. The following proposition asserts that the likelihood function is bounded above in $\overline{\Theta} \setminus S_\epsilon(S)$.

**Proposition 3.2.** For any $\epsilon > 0$, there exists a finite number $A > 0$ such that, for every sequence $\{\theta^{(m)} \in \Theta\}$ that converges to $\theta' \in \overline{\Theta} \setminus S_\epsilon(S)$, we have$$\lim_{m \to \infty} \ell^S(\theta^{(m)}) \leq A.$$**Proof:** See Appendix A.

Proposition 3.2 applies to every jump-diffusion process of (1), i.e., every distribution of $Y$. This gives us an important insight into penalizing the likelihood function. Because the unique source of singularity is the zero volatility, associating randomness to the volatility allows (numerical) searching methods to get out of the singularity point. With this in mind, we now propose a penalized likelihood function for jump-diffusion structural models,$$L_{P_\ell}^S(\theta) = L^S(\theta)p_0(\sigma),$$where $L^S(\theta)$ is the likelihood function of observed data, $L_{P_\ell}^S(\theta)$ is known as the corresponding penalized likelihood function, and $p_0(\sigma)$ which is the conditional
conjugate prior of $\sigma$ is the inverted gamma distribution with hyperparameters $a$ and $b$,

$$p_0(\sigma) = \frac{ab}{\Gamma(b-1)} \frac{1}{\sigma^{2b}} \exp \left[ -\frac{a}{\sigma^2} \right] 1_{[0, \infty)}.$$ (16)

It is clear that $p_0(\sigma)$ is bounded over $\Theta$ and that

$$\lim_{\sigma \to 0^+} p_0(\sigma)\sigma^{-N} = 0, \ \forall N.$$ 

Thus, $p_0(\sigma)$ vanishes rapidly enough to compensate for the singularities of the likelihood function. This guarantees the existence of the maximum penalized likelihood estimator and ensures that the estimator does not belong to the set of singularities. We formally write the conclusion as a proposition.

**Proposition 3.3.** For $a > 0$ and $b > 1$, the penalized likelihood function, $L^{S}_{SP}(\theta)$, is bounded above over $\Theta$. Moreover, it vanishes when $\theta$ gets close to $S_0(S)$:

$$\forall S \in (\mathbb{R}^+)^N, \theta' \in S_0(S), \lim_{\theta \to \theta'} L^{S}_{SP}(\theta) = 0.$$ 

**Proof:** See Appendix A.

Although Proposition 3.3 has shown that the inverted gamma density can be used to penalize the likelihood function, this penalty function is not unique. As our goal is to establish an estimation method for jump-diffusion structural models and due to the fact that the maximum likelihood estimator cannot be defined, the uniqueness of the penalty function is not of great interest. The advantage of using the inverted gamma density is that closed form re-estimation formulas can be derived for the MJD structural model when the estimation is implemented with an EM algorithm. This greatly reduces the computational time needed for our application.

Another concern is how the values of hyperparameters $a$ and $b$ can be chosen. Is there any guidance? We will see later that the volatility estimate is bounded below by a constant that depends on the sample size and the two hyperparameters. Therefore, $a$ and $b$ can be chosen according to the prior knowledge of the lower bound of the volatility. This point will be further clarified in Section 4.1.

The penalized likelihood function can be maximized with any numerical-search algorithm with no added programming or computational burden compared to the standard maximization of the likelihood function. Moreover, penalizing the likelihood function with the inverted gamma distribution is applicable to an
arbitrary distribution of Y. Hence, the proposed estimation is general enough to accomplish the estimation of many jump-diffusion structural models, including the MJD and KJD models. It is sometimes more convenient to maximize the penalized log-likelihood function:

$$\ell_{SP}(\theta) = \log L_{SP}(\theta) = \ell_S(\theta) - 2b \log \sigma - \frac{a}{\sigma^2} + b \log a - \log \Gamma(b - 1),$$

where $$\ell_S(\theta)$$ is defined in (13). Maximizing $$\ell_{SP}$$ over $$\Theta$$ is equivalent to maximizing:

$$\ell_S(\theta) - 2b \log \sigma - \frac{a}{\sigma^2},$$

over the same parameter space.

We now summarize the proposed estimation as follows. The estimator $$\hat{\theta}$$ is obtained by solving

$$\max_{\theta \in \Theta} \left[ \sum_{i=1}^{n} \log g(\omega_i|\omega_{i-1}, \theta) + \sum_{i=1}^{N} \left( \log \frac{S_i}{V_i} \right) - \sum_{i=1}^{N} \log \left[ \frac{\partial h}{\partial V} \right] V_i - 2b \log \sigma - \frac{a}{\sigma^2} \right]$$

subject to the constraint that,

$$S_j = h(V_j; \theta^*), \quad \forall j \in \{0, \ldots, N\},$$

where $$g(\omega_i|\omega_{i-1}, \theta)$$ is defined in (5) and $$\omega_j = \log h^{-1}(S_j, \theta^*)$$. Specifying the distribution of the jump size, Y, enables us to develop an EM algorithm to efficiently obtain the maximum penalized likelihood estimators (MPLE). EM algorithm is attractive because it produces explicit re-estimation formulas.

Slightly modifying the formulation (18) enables us to include risk-neutral parameters in the estimation. For instance, we can estimate $$(\theta, \alpha, \beta)$$ as follows,

$$\max_{\theta \in \Theta, \alpha \in [0, 1], \beta \in \mathbb{R}} \left[ \ell_S(\theta) - 2b \log \sigma - \frac{a}{\sigma^2} \right],$$

subject to the constraint that

$$S_j = h(V_j; \theta^*), \quad \forall j \in \{0, \ldots, N\},$$

where $$\theta^*$$ depends on $$\theta$$, $$\alpha$$ and $$\beta$$, under the HARA utility setting of Kou (2002). However, there is a price to pay. Firstly, the estimation quality of $$\theta$$ may be distorted by introducing two additional parameters: $$\alpha$$ and $$\beta$$. Secondly, the estimation time will be (much) longer. Finally, EM algorithm becomes very sophisticated if not impossible. For the purpose of comparing credits in a consistent manner, it may be more desirable for us to fix the values of $$\alpha$$ and $$\beta$$. From now on, we assume that $$\alpha = \beta = 1$$. This consideration is consistent with the Merton (1976) risk-neutral measure of jump-diffusion. In the remaining part of the paper, we will focus on the estimation problem of (18) and (19).
4 EM Algorithm

This section establishes EM algorithms for the proposed estimation. EM algorithm of Dempster et al. (1977, DLR) provides an iterative procedure for computing maximum likelihood estimators in a situation of incomplete data. The notion of incomplete data includes the conventional sense of missing data, but it also applies to situations where the complete data represent what would be available from a hypothetical experiment. In many cases, hidden variables can be viewed as missing data. EM algorithm can easily be extended to compute MPLE. As the name implies, the EM algorithm involves two steps: the expectation step (E-step) and the maximization step (M-step).

We highlight several difficulties in implementing (18) for motivating the development of an EM algorithm. The first is the search of local maxima over a high dimensional parameter space $\Theta$. A typical jump-diffusion process consists of 5-6 model parameters. EM algorithm provides us with re-estimation formulas that produce a sequence of estimates approaching the MPLE. The second difficulty is the computation of the delta, $\frac{\partial h}{\partial V}$, in the likelihood function (18). Usually, closed form option pricing formulas are impossible under a jump-diffusion process so that the cost of computing the delta at each re-estimation step becomes very expensive. EM algorithm avoids computing the delta. The last difficulty is the root finding procedure for solving $S = h(V, \theta^*)$, where $V = \{V_0, \ldots, V_N\}$ and $S = \{S_0, \ldots, S_N\}$, to respect the constraint of the problem. However, this is never avoidable because one of our goals is to measure the firm asset values, $V$.

Using an argument similar to that of Duan et al. (2004), we examine how the EM algorithm concept of DLR allows us to avoid computing the delta. We will then develop re-estimation formulas for the MPLE. However, the concrete formulas cannot be obtained without specifying the distribution of the jump size $Y$. To illustrate these ideas, we derive re-estimation formulas for the MJD and KJD structural models.

Because $\{\log S_i|\log S_{i-1} : i = 1, \ldots, N\}$ is a sample of i.i.d. random variables according to the model assumption, the likelihood function on log-stock prices can be thought of as the joint density of $S$. Specifically,

$$L^S(\theta) = f(S|\theta) \Rightarrow L^S_\theta(\theta) = p_0(\sigma)f(S|\theta).$$

Regarding the firm’s asset values $V$ as missing data, we have

$$p_0(\sigma)f(S|\theta) = p_0(\sigma) \int f(S, V|\theta)dV,$$
where \( f(S, V|\theta) \) is the joint density function of \( S \) and \( V \). The problem (18) has to maximize the incomplete-data penalized log-likelihood, \( \log L_{SP}(\theta) \), subject to the constraint (19). EM algorithm approaches the problem by proceeding iteratively in terms of the complete-data penalized log-likelihood, \( \log[p_0(\sigma)f(S, V|\theta)] \). As it is unobservable, it is replaced by its conditional expectation given \( S \), using the current fit for \( \theta \).

Let \( \theta^{(0)} \) be the initial value for \( \theta \) and \( \theta^{(m)} \) be the estimate of \( \theta \) in the \( m \)-th iteration of the EM algorithm. The E-step requires the calculation of:

\[
Q_1(\theta; \theta^{(m)}) = E \left\{ \log[p_0(\sigma)f(S, V|\theta)] | S, \theta^{(m)} \right\}.
\]

Given the condition that \( S = h(V; \theta^*) \) and \( \theta^* \) depends on \( \theta \), the conditional expectation is typically easy to compute as

\[
Q_1(\theta; \theta^{(m)}) = \log[p_0(\sigma)f(S, h^{-1}(S, (\theta^*)^{(m)})|\theta)]
= \log[p_0(\sigma)f(h^{-1}(S, (\theta^*)^{(m)})|\theta)]
= \sum_{i=1}^{N} \log g(\omega^m_i | \omega^m_{i-1}, \theta) + \log p_0(\sigma),
\]

where \( \omega^m_i = \log[h^{-1}(S_i, (\theta^*)^{(m)})] \) and \( g \) is the density function defined in (5).

From the last line of (21), it can be seen that \( Q_1(\theta; \theta^{(m)}) \) is essentially the penalized log-likelihood function of \( V^{(m)} = \{e^{-\lambda \Delta t} : j = 0, \ldots, N\} \). The M-step requires the maximization of \( Q_1(\theta; \theta^{(m)}) \) with respect to \( \theta \). The result, i.e., \( \theta^{(m+1)} \), is the new estimate of \( \theta \) in the current iteration. The sequence of estimates obtained in this way should approach the MPLE. However, the EM procedure does not involve the computation of the delta. Another way of presenting the iteration is to say,

\[
\theta^{(m+1)} = \arg \max_{\theta} Q_1(\theta; \theta^{(m)}),
\]

with an initial value \( \theta^{(0)} \).

To obtain re-estimation formulas, we focus on the M-step above and do the maximization with another EM algorithm. We introduce the missing data of i.i.d. random variables:

\[
C = \{c_n \in \{0, 1\}, n = 1, \ldots, N \},
\]

where “\( c_n = j \)” represents a situation in which there are \( j \) jumps arriving in the time interval \((t_{n-1}, t_n]\). Let \( \pi_0 = (1 - \lambda \Delta t) \) be the probability of no jump over a time interval with length \( \Delta t \) and \( \pi_1 = 1 - \pi_0 \) that of one jump.
The objective function in the preceding M-step can be viewed as the log of the joint density function of asset values, \( V \), with a penalty \( p_0(\sigma) \). That is,

\[
Q_1(\theta; \theta^{(m)}) = \log[p_0(\sigma) f(V^{(m)}|\theta)] = \log[p_0(\sigma) L^{(m)}(\theta)],
\]

where \( L^{(m)}(\theta) \) which is defined in (8) can be explicitly expressed as

\[
L^{(m)}(\theta) = \prod_{j=1}^{N} \left[ \pi_0 f_X(\omega_j^m|\omega_{j-1}^m, \mu, \sigma) + \pi_1 f_{X+Y}(\omega_j^m|\omega_{j-1}^m, \theta_Y) \right],
\]

where \( \omega_j^m = \log h^{-1}(S_j; (\theta^*)^{(m)}) \), \( f_X \), and \( f_{X+Y} \) are defined in (6). With the missing data \( C \), the complete data likelihood can be recognized in the formula:

\[
f(V^{(m)}|\theta) = \sum_C f(V^{(m)}, C|\theta) = \sum_C f(V^{(m)}|C, \mu, \sigma, \theta_Y) P(C|\lambda).
\]

In the notion of EM algorithm, we are interested in maximizing the conditional expectation of the complete data penalized log-likelihood function. Therefore, the E-step computes,

\[
Q_2(\theta; \theta^{(m)}) = E \{ \log[p_0(\sigma) f(V^{(m)}, C|\theta)] | V^{(m)}, \theta^{(m)} \} ,
\]

which can be decomposed into \( Q_2(\theta; \theta^{(m)}) = Q_2'(\mu, \sigma, \theta_Y; \theta^{(m)}) + Q_2''(\lambda; \theta^{(m)}) \), where,

\[
Q_2'(\mu, \sigma, \theta_Y; \theta^{(m)}) = E \{ \log[p_0(\sigma) f(V^{(m)}|C, \mu, \sigma, \theta_Y)] | V^{(m)}, \theta^{(m)} \} \quad (23)
\]

\[
Q_2''(\lambda; \theta^{(m)}) = E \left[ \log P(C|\lambda)|V^{(m)}, \theta^{(m)} \right] . \quad (24)
\]

Thus, we can separate \( \lambda \) (or \( \pi_0 \) and \( \pi_1 \)) from the other parameters in the estimation procedure under EM algorithm. The M-step is then needed to maximize \( Q_2(\theta; \theta^{(m)}) \), with respect to \( \theta \).

As \( Q_1(\theta; \theta^{(m)}) \) is the log-likelihood function of a mixture distribution, the re-estimation formula for \( \pi_0 \) derived from (24) is classic:

\[
\pi_0^{(m+1)} = \frac{1}{N} \sum_{j=1}^{N} P(c_j = 0|V^{(m)}, \theta^{(m)}),
\]

by which \( \lambda^{(m+1)} \) is retrieved from \( \pi_0^{(m+1)} = 1 - \lambda^{(m+1)} \Delta t \), where \( P(c_j = 0|V^{(m)}, \theta^{(m)}) \) can be calculated by Baye’s Theorem. We present the result here:

\[
P(c_j = 0|V^{(m)}, \theta^{(m)}) = \frac{\pi_0^{(m)} f_X(\omega_j^m|\omega_{j-1}^m, \mu^{(m)}, \sigma^{(m)})}{g(\omega_j^m|\omega_{j-1}^m, \theta^{(m)})},
\]

19
where \( g \) is given by (5). The detailed calculations of (25) and (26) can be found in Chapter 1.4.3 of McLachlan and Krishnan (1997). The expression for (23) is simple:

\[
Q'_2 = \log p_0(\sigma) + \sum_{j=1}^{N} P(c_j = 0|V^{(m)}, \theta^{(m)}) \log f_X(\omega_j^{m}|\omega_{j-1}^{m}, \mu, \sigma) \\
+ \sum_{j=1}^{N} P(c_j = 1|V^{(m)}, \theta^{(m)}) \log f_{X+Y}(\omega_j^{m}|\omega_{j-1}^{m}, \mu, \sigma, \theta_Y).
\]

It remains to maximize \( Q'_2 \), which is independent of \( \lambda \), with respect to \( \mu, \sigma \) and \( \theta_Y \). This requires us to specify the model and hence the parameter \( \theta_Y \).

### 4.1 MJD Structural Models

Despite the \( \lambda \), the MJD process consists of four other parameters: \( \mu, \sigma, k, \) and \( s \). Suppose the sampling time space is a constant, i.e., \( \Delta t_i = \Delta t \), for all \( i = 1, \ldots, N \). We consider the transformation of parameters:

\[
\mu_0 = (\mu - \sigma^2 / 2) \Delta t, \tag{28}
\]

\[
\sigma_0^2 = \sigma^2 \Delta t, \tag{29}
\]

\[
\mu_1 = \mu_0 + k, \tag{30}
\]

\[
\sigma_1^2 = \sigma_0^2 + s^2 \geq \sigma_0^2. \tag{31}
\]

On differentiating \( Q'_2 \) with respect to \( \mu_0, \mu_1, \sigma_0^2, \) and \( \sigma_1^2 \) and then setting the results to zero, we obtain the re-estimation formulas,

\[
\mu_i^{(m+1)} = \frac{1}{N\pi_i^{(m+1)}} \sum_{j=1}^{N} R_j^{(m)} \frac{\pi_i^{(m)} f_i(\omega_j^{(m)}|\omega_{j-1}^{(m)}, \theta^{(m)})}{g(\omega_j^{(m)}|\omega_{j-1}^{(m)}, \theta^{(m)})}, \tag{32}
\]

\[
\left(\hat{\sigma}_i^{(m+1)}\right)^2 = \frac{1}{2b_i + N\pi_i^{(m+1)}} \left[ 2a_i + \sum_{j=1}^{N} \left( R_j^{(m)} - \mu_i^{(m)} \right) \frac{\pi_i^{(m)} f_i(\omega_j^{(m)}|\omega_{j-1}^{(m)}, \theta^{(m)})}{g(\omega_j^{(m)}|\omega_{j-1}^{(m)}, \theta^{(m)})} \right],
\]

where \( \pi_i^{(m+1)} \) has been obtained in (25), \( g \) is given in (5), \( f_0 = f_X, f_1 = f_{X+Y}, \) \( a_i = a(1-i)\Delta t, b_i = b(1-i) \) and \( R_j^{(m)} = \omega_j^{(m)} - \omega_{j-1}^{(m)} \).

Due to the constraint of (31), the re-estimation formulas for \( \sigma_j^{(m+1)} \) should be modified to

\[
\sigma_0^{(m+1)} = \hat{\sigma}_0^{(m+1)} \quad \text{and} \quad \sigma_1^{(m+1)} = \max \left( \hat{\sigma}_0^{(m+1)}, \hat{\sigma}_1^{(m+1)} \right). \tag{33}
\]
Original parameters can then be recovered from (28) to (31). Systematically, the algorithm runs as follows.

**Algorithm 4.1.** For the MJD structural model with the structure \( S = h(V; \theta^*) \), the model parameters \( \theta = (\mu, \sigma, \lambda, k, s) \) defined in (1) and (3), can be estimated with the following steps:

- **Initiation:** Select \( \theta^{(0)} \in \Theta \) as initial values of \( \theta \).
- **E-step:** At the \((m+1)\)-th iteration, compute \( V_j^{(m)} = h^{-1}(S_j; (\theta^*)^{(m)}) \) for each \( j = 0, \ldots, N \).
- **M-step:** Compute \( \lambda^{(m+1)} \) by (25) and \((\mu_0^{(m+1)}, \mu_1^{(m+1)}, \sigma_0^{(m+1)}, \sigma_1^{(m+1)})\) by (32)-(33). Then, \( \mu^{(m+1)}, \sigma^{(m+1)}, k^{(m+1)}, s^{(m+1)} \) are solved from (28) to (31).
- **Termination:** The re-estimation procedure stops at iteration \( M \) if
  \[ \| \theta^{(M)} - \theta^{(M-1)} \|_\infty < \epsilon \]
  for some pre-specified \( \epsilon \). The estimate is then given by \( \theta^{(M)} \) while the firm values are \( \{V_j^{(M)}: j = 0, \ldots, N\} \).

It may be interesting to note that adding the penalty to the likelihood function avoids having a zero volatility as the optimal \( \sigma_0^2 > \frac{2a\Delta t}{2b+N\pi_0} > \frac{2a\Delta t}{2b+N} > 0 \) at every iteration. This provides a guideline for choosing the values of \( a \) and \( b \). Suppose we are comfortable with a lower bound \( \bar{\sigma} \) for \( \sigma \). We can solve \( a \) by setting \( \sigma^2 \) to \( \frac{2a\Delta t}{b(1+a)+N} \), where we have set \( b = 1 + a \), because it is required that \( a > 0 \) and \( b > 1 \). Moreover, the value of \( b \) has a little impact on the estimation, if it is larger than 1, because the lower bound of the volatility is mainly controlled by the value of \( a \) for a large sample size.

### 4.2 KJD Structural Models

The KJD process consists of five parameters, \( \mu, \sigma, \eta_1, \eta_2, \) and \( p \), in addition to the jump arrival rate \( \lambda \). It is expected that the estimation is more sophisticated. From (7), we observe that \( f_{X+Y} \) is actually a mixture of two distributions,

\[
f_{X+Y} = pf_u + (1-p)f_d,
\]

with \( f_u \) accounting for the upward jump and \( f_d \) for the downward jump. When a jump arrives, the probability of jumping upward is \( p \) and downward \( 1 - p \).
Therefore, the re-estimation formula for $p$ can be obtained by iterating a formula similar to that of $\lambda$ (or $\pi_0$).

In addition to $C$, we introduce the following hypothetical missing data,

$$D = \{d_j \in \{1, 2\} : j \in \{1, \ldots, N\}\},$$

where “$d_j = 1$” (“$d_j = 2$”) represents an asset price that jumps upwardly (downwardly) at the time interval $(t_{j-1}, t_j]$. Hence, we define the probabilities, $\pi_{11} = \pi_1 p$ and $\pi_{12} = \pi_1 (1 - p)$, where $\pi_1 = \lambda \Delta t$ for a constant sampling time space $\Delta t$.

The complete data penalized log-likelihood becomes,

$$Q_3(\theta; \theta^{(m)}) = E\{\log[p_0(\sigma)f(V^{(m)}, C, D|\theta)]|V^{(m)}, \theta^{(m)}]\},$$

which can be decomposed into $Q_3 = Q'_3 + Q''_3$, where

$$Q'_3(\mu, \sigma, \eta_1, \eta_2; \theta^{(m)}) = E\{\log[p_0(\sigma)f(V^{(m)}|C, D, \mu, \sigma, \eta_1, \eta_2)]|V^{(m)}, \theta^{(m)}]\},$$
$$Q''_3(\lambda, p; \theta^{(m)}) = E[\log P(C, D|\lambda, p)|V^{(m)}, \theta^{(m)}].$$

Using standard re-estimation formulas for the weighting probabilities of a mixture distribution, (McLachlan and Krishnan, 1997), we find that:

$$\pi_{0}^{(m+1)} = \frac{1}{N} \sum_{j=1}^{N} P(c_jd_j = 0|V^{(m)}, \theta^{(m)}) := \frac{1}{N} \sum_{j=1}^{N} P_{0j},$$

(37)

$$\pi_{11}^{(m+1)} = \frac{1}{N} \sum_{j=1}^{N} P(c_jd_j = 1|V^{(m)}, \theta^{(m)}) := \frac{1}{N} \sum_{j=1}^{N} P_{1j},$$

(38)

$$\pi_{12}^{(m+1)} = \frac{1}{N} \sum_{j=1}^{N} P(c_jd_j = 2|V^{(m)}, \theta^{(m)}) := \frac{1}{N} \sum_{j=1}^{N} P_{2j}.$$  

(39)

Obviously, $P(c_j = 0) = P(c_jd_j = 0)$ so that the formula for $\pi_0$ in (37) is exactly the same as in (25). By Baye’s theorem, we obtain that,

$$P_{1j} = \frac{\pi_{11}^{(m)} f_{u}(\omega_{j}^{m}|\omega_{j-1}^{m}, \theta^{(m)})}{g(\omega_{j}^{m}|\omega_{j-1}^{m}, \theta^{(m)})},$$
$$P_{2j} = \frac{\pi_{12}^{(m)} f_{d}(\omega_{j}^{m}|\omega_{j-1}^{m}, \theta^{(m)})}{g(\omega_{j}^{m}|\omega_{j-1}^{m}, \theta^{(m)})},$$

where $f_{u}, f_{d}$ are defined in (34) and $g$ is given by (5). After calculating $\lambda$ from (37), the estimate of $p$ can be obtained from $\pi_{11} = \lambda \Delta t p$ through (38).
It remains to determine \( \mu, \sigma, \eta_1, \) and \( \eta_2 \) by maximizing:

\[
Q_3' (\mu, \sigma, \eta_1, \eta_2; \theta^{(m)}) = \log p_0 (\sigma) + \sum_{j=1}^{N} P_{0j} \log f_X (\omega_j^m | \omega_{j-1}^m, \mu, \sigma) + \sum_{j=1}^{N} P_{1j} \log f_u (\omega_j^m | \omega_{j-1}^m, \mu, \sigma, \eta_1, \eta_2) + \sum_{j=1}^{N} P_{2j} \log f_d (\omega_j^m | \omega_{j-1}^m, \mu, \sigma, \eta_1, \eta_2) \tag{40}
\]

Unfortunately, setting the partial derivatives of \( Q_3' \) to zero produces a system of non-linear equations that does not admit closed form re-estimation formulas. See appendix B.

When it is infeasible to attempt to find the maxima in closed form, DLR defines a generalized EM (GEM) algorithm in which the M-step requires \( \theta^{(m+1)} \) to be chosen, such that

\[
Q_3'(\theta^{(m+1)}; \theta^{(m)}) \geq Q_3'(\theta^{(m)}; \theta^{(m)}) \tag{41}
\]

holds. That is, one chooses \( \theta^{(m+1)} \) to increase \( Q_3'(\theta; \theta^{(m)}) \), rather than to maximize it over \( \Theta \). It has been shown that the sequence of GEM iterates converges to a stationary point. One way to construct a GEM algorithm is based on one Newton-Raphson step.

Let \( \psi = (\psi_1, \psi_2, \psi_3, \psi_4) = (\mu, \sigma^2, \eta_1, \eta_2) \in \mathbb{R}^{4 \times 1} \) be the vector collecting the four remaining parameters. The GEM algorithm runs as follows:

\[
\psi^{(m+1)} = \psi^{(m)} - c^{(m)} H_\psi (\psi^{(m)}; \theta^{(m)})^{-1} J_\psi (\psi^{(m)}; \theta^{(m)}) , \tag{42}
\]

where

\[
J_\psi (\psi; \theta^{(m)}) = \left[ \frac{\partial Q_3'(\psi; \theta^{(m)})}{\partial \psi_i} \right] \in \mathbb{R}^{4 \times 1} ,
\]

\[
H_\psi (\psi; \theta^{(m)}) = \left[ \frac{\partial^2 Q_3'(\psi; \theta^{(m)})}{\partial \psi_i \partial \psi_j} \right] \in \mathbb{R}^{4 \times 4} , \tag{43}
\]

and \( c^{(m)} > 0 \) is chosen, such that (41) holds. \( H_\psi (\psi^{(m)}; \theta^{(m)}) \) and \( J_\psi (\psi^{(m)}; \theta^{(m)}) \) are respectively known as the Hessian and Jacobian of \( Q_3' \), and their formulas appear in Appendix B. It can be seen that (42) is the same as the first step in the
standard Newton-Raphson method if \( c^{(m)} \equiv 1 \). Moreover, (41) must hold if \( c^{(m)} \) is sufficiently close to zero. Thus, the increase in \( Q_3' \) is always granted with a small \( c^{(m)} \). One can start with \( c^{(m)} = 1 \) for each iteration. If (41) is satisfied, we proceed to the next step. Otherwise, we try another \( c^{(m)} \) by dividing the previous one by 2 until (41) is satisfied. We must always keep in mind that \( \lambda^{(m+1)} \) and \( p^{(m+1)} \) should be obtained in advance. It may be useful to summarize the whole algorithm.

Algorithm 4.2. For the KJD structural model with the structure, \( S = h(V; \theta^*) \), the model parameters \( \theta = (\mu, \sigma, \lambda, p, \eta^1, \eta^2) \) defined in (1) and (4), can be estimated using the same steps as in Algorithm 4.1, except that the M-step is replaced:

- Compute \( \lambda^{(m+1)} \) by (25), \( p^{(m+1)} \) by (38), and \( (\mu^{(m+1)}, \sigma^{(m+1)}, \eta_1^{(m+1)}, \eta_2^{(m+1)}) \) through (42)-(43).

Remarks:

1. In fact, the GEM algorithm can be extended to obtain the MPLE for all jump-diffusion processes through the formula of \( Q'_2 \) in (27). This requires the computation of the Jacobian and Hessian of \( Q'_2 \). Hence, our approach is not limited to MJD and KJD processes.

2. The EM algorithm for MJD structural models is attractive because the closed form re-estimation formulas for all parameters are available so that the computational time can be reduced greatly.

3. The GEM algorithm for KJD models is also attractive because the closed form re-estimation formula for \( p \) is obtained. This reduces the dimension of the Hessian matrix used in the GEM algorithm.

4.3 Computation of the E-step

The E-step requires the calculation of the inverse of an option pricing formula at each sampling time point and repeats the whole calculation per iteration. For example, a typical estimation in the market is based on daily stock prices over a year. If there are 250 trading days in the year, then the estimation requires the computation of the inverse for 250 times per iteration. Thus, the computation of the E-step is of a great concern in practice. If the closed form pricing formula is available and its computational speed is fast, for example, less than 0.1 second for
pricing an option, then performing the standard numerical root-search method is an acceptably efficient way.

However, as we have seen in Section 2, closed form solutions are not always possible. For example, the DOC option price under the MJD process should rely on simulation or other numerical methods. Although analytical solutions exist for some special cases, the computation of the solution may not be efficient enough to implement the EM algorithm to a practical standard. For example, option pricing under the KJD process involves a Laplace inversion or even a double Laplace inversion.

To effectively reduce the estimation time, we propose a modified E-step based on one quasi-Newton step. This idea is borrowed from the GEM algorithm in which the M-step is based on one Newton-Raphson step, but we are considering the E-step now. A quasi-Newton method is used to find the unique root of a nonlinear algebraic equation. In our case, we are interested in solving:

\[ H_m(V) = S - h(V, \theta^{(m)}) = 0. \] (44)

The standard Newton method considers the fixed point iteration,

\[ V^{(j+1)} = V^{(j)} - H_m \left( V^{(j)} \right) \left[ \frac{dH_m}{dV} \right]^{-1}_{V=V^{(j)}}, \]

which requires us to compute the option delta. The modified E-step uses the first step of a quasi-Newton method, in which the derivative in the standard Newton method is approximated by a finite difference. We propose that,

\[ V^{(m+1)} = V^{(m)} - \frac{H_{m+1} \left( V^{(m)} \right) \left( V^{(m)} - \tilde{V}^{(m)} \right)}{H_{m+1} \left( V^{(m)} \right) - H_{m+1} \left( \tilde{V}^{(m)} \right)} \]

\[ \tilde{V}^{(m+1)} = \begin{cases} V^{(m)}, & \text{if } H_{m+1} \left( V^{(m+1)} \right) H_{m+1} \left( V^{(m)} \right) < 0; \\ \tilde{V}^{(m)}, & \text{otherwise if } H_{m+1} \left( V^{(m+1)} \right) H_{m+1} \left( \tilde{V}^{(m)} \right) < 0; \\ V^{(0)}, & \text{otherwise if } H_{m+1} \left( V^{(m+1)} \right) < 0; \\ \tilde{V}^{(0)}, & \text{otherwise}, \end{cases} \] (46)

where \( V^{(0)} \) and \( \tilde{V}^{(0)} \) are chosen, such that \( H_m(V^{(0)}) > 0 \) and \( H_m(\tilde{V}^{(0)}) < 0 \) for all \( m \). For our application, \( V^{(0)} = S \) and \( \tilde{V}^{(0)} = S + K \) are used because they are the lower and upper bound of the true asset value. It is ensured by (46) that
the root always lies in between \( V^{(m)} \) and \( \bar{V}^{(m)} \). Our simulation shows that this proposal is very efficient and does not produce additional error in the estimator.

The finite difference approximation in (45) may be unstable when simulation is used to compute the equity value. To overcome this, we stall all of the generated random numbers for computing the option price in the first iteration and then reuse them in all of the remaining iterations. Hence, the estimation is subject to the bias of the generated sample. The solution is to base the simulation on many asset price paths and on some variance reduction techniques.

5 Performance of Estimation

This section examines the accuracy and efficiency of the ML estimation (13) implemented with a direct numerical-search method and of the PLE (18) with EM algorithm. Programs are coded with MATLAB, and the direct numerical-search refers to the routine “fminsearch” provided by the software. We apply the aforementioned approaches to the MJD and KJD structural models. The capital structure of both barrier-independent (Merton) and barrier-dependent (Black-Cox) models are considered. Table 1 presents abbreviations of the approaches.

<table>
<thead>
<tr>
<th>Table 1: Abbreviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML estimation (MLE)</td>
</tr>
<tr>
<td>PLE with EM algorithm (PEM)</td>
</tr>
</tbody>
</table>

The comparison is two-fold: using simulation and empirical analysis. In the simulation study, we perform the estimation in a hypothetical environment that agrees perfectly with the model assumption. We check whether the proposed estimation produces estimates consistent with the input parameters. This enables us to comment on the accuracy of each approach. We also compare the execution times.

We should clarify here that the empirical study is not used to check the performance of jump-diffusion structural models in predicting credit spread or PD. Instead, supposing a model has been chosen in practice, we are interested in the consequences of implementing the model with different estimation methods. Our empirical study shows that the ML estimation requires a relatively long computational time and sometimes produces a zero volatility estimate. In contrast, the
The proposed estimation method is very efficient and the volatility estimate is always positive.

We also compare the diffusion process and the MJD process with empirical data. We attempt to show empirically that the estimated firm asset value and its volatility can be very different in the two situations. This implies that it may be inappropriate to estimate firm value and volatility under a diffusion process in advance and then fit the jump component to the residual credit spread. Therefore, the proposed estimation is an indispensable tool for implementing jump-diffusion structural credit risk models. The empirical performance of the MJD and KJD structural models in predicting credit spreads will, however, be reported in a separate paper because it involves empirical considerations beyond the scope of this paper.

5.1 Monte Carlo Evidence

<table>
<thead>
<tr>
<th></th>
<th>MJD process</th>
<th>KJD process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asset value</td>
<td>$V(0) = 100$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu = 15%$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma = 25%$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\lambda = 10$ per annum</td>
<td></td>
</tr>
<tr>
<td>Jump Size</td>
<td>$k = -5%$</td>
<td>$p = 0.6$</td>
</tr>
<tr>
<td>Parameters</td>
<td>$s = 10%$</td>
<td>$\eta_1 = \eta_2 = 20$</td>
</tr>
<tr>
<td>Option Interest rate</td>
<td>$r = 5%$</td>
<td></td>
</tr>
<tr>
<td>Pricing Asset payout ratio</td>
<td>$q = 0$</td>
<td></td>
</tr>
<tr>
<td>Parameters Option maturity</td>
<td>$T = 3$ years</td>
<td></td>
</tr>
<tr>
<td>Promised payment</td>
<td>$K = 50$</td>
<td></td>
</tr>
</tbody>
</table>

We simulate the firm asset value process (2) under the physical probability measure. Sample paths are generated using the Euler method with a time-space of $1/250$, replicating daily asset prices over a year. Both MJD and KJD processes are considered. At each time point, the asset value is transformed into a stock price based on the chosen structural model, i.e., the corresponding option pricing formula (see Section 2). For the barrier-dependent MJD model, we use the
Brownian Bridge simulation of Metwally and Atiya (2002) to calculate the DOC option price, as no analytical formula is available. The Brownian Bridge simulation is based on 10,000 asset price paths and 10,000 antithetic paths under the risk-neutral measure. Parameter values for the simulation are summarized in Table 2.

We discard the asset price paths and regard the generated stock prices as observed data in the estimation procedure. To examine the quality of estimation, simulation and estimation are repeated for 100 times. The estimation uses the hyperparameters: $a = 0.4$ and $b = 1.4$. Table 3 and 4 display the results of MJD and KJD processes, respectively. In both tables, the firm value error for each sample path is calculated as:

$$\text{Firm value error} = \frac{1}{N} \sum_{j=1}^{N} \frac{\hat{V}_j - V_j}{V_j},$$

where $\hat{V}_j$ is the estimated firm value at time $t_j$, $V_j$ is the true asset value and $N$ is the total number of time points. Moreover, the number of ill cases is also reported. An ill case represents a situation in which the estimation program keeps running over 24 hours and should be manually terminated.

We concentrate on the MJD structural models first. From Table 3, we see that there are 35 ill cases for the barrier-independent model and 100 for the barrier-dependent model with the ML estimation. This means that the ML estimation would give no output for the MJD barrier-dependent model after running the estimation program for a whole day. In fact, the ML estimation can produce an output for some particular sample paths after running the program over two days. However, the numbers are completely unreliable, and we discontinue the process due to the unreasonably long computational time. The failure rate of the ML estimation is around 35% for the barrier-independent model. After eliminating ill cases from the report, the ML estimation requires an average execution time of 5,580 seconds, which is equivalent to around 1.5 hours per sample path, and the reported estimates are significantly biased upward for the barrier-independent model.

In contrast, the penalized likelihood methods produce no ill cases and render accurate estimates. In particular, PEM is very efficient in the sense that it only takes 16 seconds on average for estimating the barrier-independent model and 4,458 seconds for the barrier-dependent model. It is seen that the execution time can be significantly reduced by utilizing EM algorithm and the analytical pricing formula. The estimates are close to the corresponding true values and the standard derivations are small, except for the estimate of $\mu$. It is well known that the drift
Table 3: MJD structural models

<table>
<thead>
<tr>
<th>True value</th>
<th>Estimated value</th>
<th>Barrier-Independent</th>
<th>Barrier-Dependent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MLE</td>
<td>PEM</td>
</tr>
<tr>
<td>( \mu ) (0.15)</td>
<td>0.542</td>
<td>0.142</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(1.405)</td>
<td>(0.394)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>( \sigma ) (0.25)</td>
<td>0.327</td>
<td>0.25</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.251)</td>
<td>(0.016)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>( \lambda ) (10)</td>
<td>40.66</td>
<td>10.25</td>
<td>10.47</td>
</tr>
<tr>
<td></td>
<td>(45.21)</td>
<td>(5.06)</td>
<td>(5.53)</td>
</tr>
<tr>
<td>( k ) (-0.05)</td>
<td>0.24</td>
<td>-0.055</td>
<td>-0.049</td>
</tr>
<tr>
<td></td>
<td>(1.925)</td>
<td>(0.052)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>( s ) (0.1)</td>
<td>0.21</td>
<td>0.074</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>(0.332)</td>
<td>(0.030)</td>
<td>(0.0285)</td>
</tr>
<tr>
<td>Firm value error (0)</td>
<td>-3.49%</td>
<td>0.05%</td>
<td>-0.4%</td>
</tr>
<tr>
<td></td>
<td>(17.97%)</td>
<td>(1.51%)</td>
<td>(4.35%)</td>
</tr>
<tr>
<td>Number of ill cases</td>
<td>35</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Execution time</td>
<td>5580</td>
<td>15.85</td>
<td>&gt;86400</td>
</tr>
<tr>
<td></td>
<td>(1448)</td>
<td>(2.73)</td>
<td>(1398)</td>
</tr>
</tbody>
</table>

requires a long time series for an accurate estimate. Duan (1994) reports that \( \mu \) is difficult to estimate correctly even under the diffusion process. Fortunately, the drift is useless for pricing credit related instruments and for calculating PD under the Merton risk-neutral measure. As expected, the estimation errors of \( k \) and \( s \) are relatively large compared to those of \( \sigma \). As we use \( \lambda = 10 \), there are only 10 jumps on average in each sample path. The estimation of jump parameters \( k \) and \( s \) is essentially based on an average of 10 jumps. In this sense, the estimation quality is acceptable. To improve the accuracy in practice, one may consider a long time series or tick-to-tick stock data.

The proposed estimation also works well for the KJD structural models and obviously outperforms the ML estimation. We see from Table 4 that the ML estimation produces 28 ill cases for the barrier-independent model and no output for the barrier-dependent model within one day. The estimation error is huge. In contrast, PEM demonstrates its clear advantage in this simulation exercise. Although
Table 4: KJD structural models

<table>
<thead>
<tr>
<th>True value</th>
<th>Estimated value</th>
<th>Barrier-Dependent</th>
<th>Barrier-Dependent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MLE</td>
<td>PEM</td>
</tr>
<tr>
<td>$\mu$ (0.15)</td>
<td>0.6442</td>
<td>0.1635</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(1.7324)</td>
<td>(0.2611)</td>
<td>(0.2931)</td>
</tr>
<tr>
<td>$\sigma$ (0.25)</td>
<td>0.311</td>
<td>0.2509</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.2662)</td>
<td>(0.0205)</td>
<td>(0.0128)</td>
</tr>
<tr>
<td>$\lambda$ (10)</td>
<td>37.743</td>
<td>10.0962</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(47.855)</td>
<td>(6.6379)</td>
<td>(6.5709)</td>
</tr>
<tr>
<td>$p$ (0.6)</td>
<td>0.8668</td>
<td>0.5924</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.5026)</td>
<td>(0.319)</td>
<td>(0.3035)</td>
</tr>
<tr>
<td>$\eta_1$ (20)</td>
<td>9.345</td>
<td>20.5366</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(13.898)</td>
<td>(11.817)</td>
<td>(11.2619)</td>
</tr>
<tr>
<td>$\eta_2$ (20)</td>
<td>10.6453</td>
<td>29.1178</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(21.6884)</td>
<td>(17.0667)</td>
<td>(15.291)</td>
</tr>
<tr>
<td>Firm value error (0)</td>
<td>-6.44%</td>
<td>-0.8157%</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(19.783%)</td>
<td>(1.7145%)</td>
<td>(1.8223%)</td>
</tr>
<tr>
<td>Number of ill cases</td>
<td>28</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Execution time</td>
<td>9400</td>
<td>433</td>
<td>&gt;86400</td>
</tr>
<tr>
<td></td>
<td>(5693)</td>
<td>(156)</td>
<td>(423)</td>
</tr>
</tbody>
</table>

As the number of parameters increases, the PEM still provides fairly good estimates within a reasonable computational time. The estimate of $\eta_2$ contains a relatively large error because both $\lambda$ and $1 - p$ are small. When $\lambda$ is set to 10 and $p$ to 0.6, we have 10 jumps on average for each sample path and four of them are downward jumps. The average downward jump size $\eta_2$ is essentially estimated from four observations on average. Thus, the large estimation error of $\eta_2$ is unavoidable for all estimation method if both $\lambda$ and $1 - p$ are small values.

To further illustrate the effect of $\lambda$, Table 5 reports the estimates for the MJD and KJD barrier-independent models when $\lambda = 20$. It can be seen that the quality of the estimation is improved for the jump component. In particular, the average estimation error of $s$ reduces from -0.026 to -0.005 and the estimation error of $\eta_2$ from 9.12 to 0.76. Meanwhile, the quality of estimation for the diffusion...
Table 5: Estimation with PEM: $\lambda = 20$

<table>
<thead>
<tr>
<th></th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\lambda$</th>
<th>$k$</th>
<th>$s$</th>
<th>$p$</th>
<th>$\eta_1$</th>
<th>$\eta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Value</td>
<td>0.15</td>
<td>0.25</td>
<td>20</td>
<td>-0.05</td>
<td>0.1</td>
<td>0.6</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>MJD</td>
<td>0.143</td>
<td>0.251</td>
<td>21.11</td>
<td>-0.053</td>
<td>0.095</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.401)</td>
<td>(0.018)</td>
<td>(6.924)</td>
<td>(0.049)</td>
<td>(0.021)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KJD</td>
<td>0.170</td>
<td>0.252</td>
<td>19.12</td>
<td>—</td>
<td>—</td>
<td>0.567</td>
<td>19.09</td>
<td>20.76</td>
</tr>
<tr>
<td></td>
<td>(0.271)</td>
<td>(0.015)</td>
<td>(8.107)</td>
<td>—</td>
<td>—</td>
<td>(0.188)</td>
<td>(6.713)</td>
<td>(9.982)</td>
</tr>
</tbody>
</table>

component remains almost the same.

5.2 Empirical Evidence

This section presents an empirical investigation of the MJD-Merton model, which is applicable to firms that have simple capital structures. Following Eom et al. (2004) and Li and Wong (2006), we consider firms with simple capital structures as firms that have only one or two public bonds and the bonds are not sinkable or subordinated bonds. The characteristics of the firms are examined with information provided by the Rating Interactive of Moody’s Investor Services. We regard firms with an organization type as corporations and exclude those with a non-US domicile. Firms in broad industries such as finance, real estate finance, public utility, insurance and banking are also excluded from our sample. At this stage, our sample consists of 2,033 firms.

To measure the market value of corporate assets, we restrict our attention to firms that have issued equity and provide regular financial statements. Therefore, we downloaded the market values of equities in the period of 1986 to 1996 from Datastream. This period of time is also used by Eom et al. (2004) and Li and Wong (2006).

Table 6 presents the basic statistics of our data. It exhibits the mean values of Moody’s rating, S&P rating, market capitalizations, and total liabilities across years. The means of the Moody’s and S&P ratings are calculated by assigning a number to a rating class. For the Moody’s rating, 1 stands for Aaa+, 2 stands for Aaa, and so on. For the S&P ratings, 1 stands for AAA+, 2 stands for AAA, and so on. For both rating systems, 24 stands for NR, which means that the bond is not
rated. The annualized risk-free rate, shown in the Panel B, ranges from the lowest of 1.32% to the highest of 5.98% in the period. The mean value is 4.3%. In this paper, we use constant interest rates for each year. The rate of a year is calculated as the average annual interest rate over the year. We ignore the dividend and stock repurchase in this empirical study.

Table 6: Descriptive statistics for the sample

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of Firms</th>
<th>Moody’s ratings</th>
<th>S&amp;P ratings</th>
<th>Average Equity Value</th>
<th>Average Total Debt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1986</td>
<td>46</td>
<td>6.95</td>
<td>6.65</td>
<td>4479.68</td>
<td>4622.74</td>
</tr>
<tr>
<td>1987</td>
<td>58</td>
<td>5.93</td>
<td>5.93</td>
<td>6309.20</td>
<td>5575.82</td>
</tr>
<tr>
<td>1988</td>
<td>76</td>
<td>6.45</td>
<td>6.26</td>
<td>5286.23</td>
<td>9584.63</td>
</tr>
<tr>
<td>1989</td>
<td>77</td>
<td>6.69</td>
<td>6.46</td>
<td>6355.56</td>
<td>8661.61</td>
</tr>
<tr>
<td>1990</td>
<td>80</td>
<td>6.31</td>
<td>6.27</td>
<td>8371.26</td>
<td>10086.37</td>
</tr>
<tr>
<td>1991</td>
<td>105</td>
<td>6.46</td>
<td>6.25</td>
<td>8573.71</td>
<td>5124.63</td>
</tr>
<tr>
<td>1992</td>
<td>116</td>
<td>7.25</td>
<td>6.77</td>
<td>6892.26</td>
<td>4050.76</td>
</tr>
<tr>
<td>1993</td>
<td>137</td>
<td>7.30</td>
<td>6.90</td>
<td>7572.97</td>
<td>4120.06</td>
</tr>
<tr>
<td>1994</td>
<td>152</td>
<td>7.55</td>
<td>7.17</td>
<td>7752.24</td>
<td>4518.75</td>
</tr>
<tr>
<td>1995</td>
<td>192</td>
<td>7.66</td>
<td>7.43</td>
<td>8107.86</td>
<td>4203.99</td>
</tr>
<tr>
<td>1996</td>
<td>197</td>
<td>8.07</td>
<td>7.95</td>
<td>7754.13</td>
<td>3231.63</td>
</tr>
</tbody>
</table>

Table B: Risk free rate

<table>
<thead>
<tr>
<th>Mean</th>
<th>Standard Derivation</th>
<th>Minimum</th>
<th>Median</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0428</td>
<td>0.0148</td>
<td>0.0132</td>
<td>0.0486</td>
<td>0.0598</td>
</tr>
</tbody>
</table>

We estimate parameters for the MJD barrier-independent structural model, i.e., the Merton (1974) model under the Merton (1976) jump-diffusion process. Table 7 displays the estimation result by year. It can be seen that the impact of jumps is significant. The jump arrival rate ranges from 7 to 15 over the years. These figures indicate that the jump component plays a substantial role in credit risk models. Although it is not included in the table, we also examine the ML estimation for the MJD model. The execution time is almost a hundred times of that of the EM algorithm using our sample. Around 20% of the sample paths are ill cases.

Many papers have examined the jump effects on major financial indices and document that the jump arrival rates are usually fewer than six per year. We, how-
ever, show that the arrival rate is much larger for an individual firm’s assets. The reason for this is that major financial index is a portfolio of many individual stocks in which the diversification may have already diluted the jump effect. When a stock price jumps up, there may be another stock price jumps down to compensate the jump effect on the financial index. Credit risk, which is a firm-specific risk, measures individual companies’ default probability. Thus, estimation should be done in a firm-specific manner. Our approach provides a possibility for this purpose.

As there was no estimation method for jump-diffusion structural models available before, empirical studies have estimated the drift and volatility using a diffusion model and then calibrated the jump distribution to the residual credit spread. Traditionally, the jump arrival rate is specified, rather than estimated, such that it is consistent with that of a major financial index. We see that this may under-estimate the arrival rate significantly. Besides, the estimation of the firm value and the volatility produced by the diffusion model may be inconsistent with those produced by jump-diffusion models. Table 8 shows this effect. After the firm value and volatility are estimated using both the diffusion model and the MJD model, we can compare the percentage differences between the two approaches. It can be seen that the percentage difference is quite large. Although the volatility estimate is larger in the diffusion model, the firm value estimate is smaller.

6 Conclusion

We have investigated the estimation of jump-diffusion structural credit risk models in this paper. It has been shown that the maximum likelihood estimator cannot be defined because the likelihood function for the equity return may blow up to infinity. We have identified the set of singularity points and proved that the likelihood function is bounded outside any neighborhood of the singularity set. We then proposed a penalized likelihood estimation in which the likelihood function is penalized with a prior distribution on the volatility, the unique source of singularity. It has been proven that the penalized likelihood function is always bounded and the maximum penalized likelihood estimator (MPLE) is well defined.

To improve the efficiency, we further established EM algorithms to iteratively obtain the MPLE under different situations. Although a generalized EM algorithm can be applied to general jump-diffusion processes, we focused on the Merton jump-diffusion (MJD) and Kou jump-diffusion (KJD) processes for the purpose of illustration. In particular, closed form re-estimation formulas for the MJD
structural models and semi-closed form re-estimation formulas for the KJD model have been obtained. It is supported by simulations that the proposed estimation is efficient and accurate.

We have also provided the first empirical evidence that the jump component is a significant factor in the firm value process under the Merton structural framework. Future research may consider the performance of jump-diffusion structural models in pricing corporate bonds using the proposed estimation method.
Appendices

A Proofs

A.1 Proof of Proposition 3.1

In the same manner of \( \theta \), we define \( \theta^* = (\theta^*_X, \theta^*_Y) \) and \( \theta' = (\theta'_X, \theta'_Y) \), where \( \theta^* \) is the vector of risk-neutral parameters and \( \theta' \in \overline{\Theta}_X \times \Theta_Y \). Under the HARA utility, \( \theta^* \) depends on \( \theta, \alpha \) and \( \beta \), but not on \( \mu \). Thus, we also use the notation: \( (\theta^*)' \) and \( (\theta^*)^{(m)} \), to indicate the dependence of \( \theta^* \) on \( \theta' \) and \( \theta^{(m)} \), respectively.

The principle of the proof is to construct a particular sequence \( \{ \theta^{(m)} \in \Theta \} \) that satisfies the statement of Proposition 3.1.

Since \( \theta' \in S_0(S), \exists n \in \{1, \ldots, N \} \) such that \( \mu' \Delta t = \log(V_n' / V_{n-1}') \) where \( S_n = h(V_n', (\theta^*)') \) and \( S_{n-1} = h(V_{n-1}', (\theta^*)') \). The required sequence \( \theta^{(m)} = (\theta_X^{(m)}), \theta_Y^{(m)}) \) can be constructed as follows:

1. \( \forall m, \theta_Y^{(m)} = \theta_Y' \);  
2. \( \lambda^{(m)} \Delta t = (1 - 1/m) \lambda' \Delta t + 1/(2m) \);  
3. \( \sigma^{(m)} = e^{-m} \);  
4. \( V_j^{(m)} = h^{-1}(S_j; (\theta^*)^{(m)}) \);  
5. \( \mu^{(m)} \Delta t = \log(V_n^{(m)} / V_{n-1}^{(m)}) + (\sigma^{(m)})^2 \Delta t/2. \)

It is clear that \( \theta^{(m)} \in \Theta \) and \( \theta^{(m)} \to \theta' \in S_0(S) \). Define \( \pi_0^{(m)} = 1 - \lambda^{(m)} \Delta t \), \( \pi_1^{(m)} = \lambda^{(m)} \Delta t \) and \( \omega_j^{(m)} = \log V_j^{(m)} \). Then, we have \( \pi_j^{(m)} \geq 1/(2m) \) and

\[
L^S(\theta^{(m)}) = \prod_{j=1}^N \left( \pi_0^{(m)} f_X(\omega_j^{(m)} | \omega_{j-1}^{(m)}, \theta^{(m)}) + \pi_1^{(m)} f_X + Y(\omega_j^{(m)} | \omega_{j-1}^{(m)}, \theta^{(m)}) \right) \left[ \frac{S_j}{V} \left( \frac{\partial h}{\partial V} \right) \right]^{-1} \theta^* = (\theta^*)^{(m)} \]

\[
\geq 1 \left( \frac{1}{2m} \right)^N \prod_{j=1}^N \left( f_X(\omega_j^{(m)} | \omega_{j-1}^{(m)}, \theta^{(m)}) + f_X + Y(\omega_j^{(m)} | \omega_{j-1}^{(m)}, \theta^{(m)}) \right) \left[ \frac{S_j}{V} \left( \frac{\partial h}{\partial V} \right) \right]^{-1} \theta^* = (\theta^*)^{(m)} .
\]

We now investigate lower bounds for each of the \( N \) terms of the product. For the moment, we set aside the elasticity ratio, i.e., the term in the square bracket above, and concentrate on the lower bounds of the remaining terms.
• For \( j = n \),
\[
f_X(\omega_j^{(m)}|\omega_{j-1}^{(m)}, \theta^{(m)}) + f_X + Y(\omega_j^{(m)}|\omega_{j-1}^{(m)}, \theta^{(m)}) \geq \frac{1}{\sqrt{2\pi}\sigma^{(m)}}.
\]
• For all \( j \neq n \),
\[
f_X(\omega_j^{(m)}|\omega_{j-1}^{(m)}, \theta^{(m)}) + f_X + Y(\omega_j^{(m)}|\omega_{j-1}^{(m)}, \theta^{(m)}) \geq f_X + Y(\omega_j^{(m)}|\omega_{j-1}^{(m)}, \theta^{(m)}).
\]

From (6), we have:
\[
f_{X+Y}(\omega_j^{(m)}|\omega_{j-1}^{(m)}, \theta^{(m)}) = \int_{-\infty}^{\infty} f_X(y - \omega_i|\omega_{i-1}, \theta_X^{(m)}) f_{Y}(y|\theta_Y^{(m)}) dy
\]
\[
\rightarrow \int_{-\infty}^{\infty} \delta(y - (\omega_i - \omega_{i-1} - \mu't \Delta t)) f_{Y}(y|\theta_Y^{(m)}) dy
\]
\[
= f_Y(\omega_i - \omega_{i-1} - \mu't \Delta t|\theta_Y^{(m)}) > 0. \tag{47}
\]

Hence, we know that:
\[
L^S(\theta^{(m)}) \geq \prod_{j=1}^{N} \frac{\left[ S_j V_j \left( \frac{\partial h}{\partial V} \right)^{-1} \right]^{\theta^*=(\theta^*)^{(m)}}}_{V=V_j^{(m)}} \prod_{j \neq n} \int_{-\infty}^{\infty} f_X(y - \omega_i|\omega_{i-1}, \theta_X^{(m)}) f_{Y}(y|\theta_Y^{(m)}) dy.
\]

For general jump-diffusion processes, Bergman et al. (1996) show that the delta of European contingent claims is bounded and hence \((\partial h/\partial V)^{-1}\) is bounded away from zero. Moreover, the limit in (47) guarantees that the integration is bounded away from zero. Hence, the lower bound of \(L^S(\theta^{(m)})\) diverges to infinity through the sequence of \(\sigma^{(m)} = e^{-m}\). This completes the proof.

**A.2 Proof of Proposition 3.2**

When \( \sigma > \epsilon \), we have:
\[
f_X(\omega_j|\omega_{j-1}, \theta) \leq \left( \epsilon \sqrt{2\pi} \right)^{-1};
\]
\[
f_{X+Y}(\omega_j|\omega_{j-1}, \theta) = \int_{-\infty}^{\infty} f_Y(\omega_j - y|\omega_{j-1}, \theta_Y) f_X(y|\omega_{j-1}, \theta_X) dy
\]
\[
\leq \left( \epsilon \sqrt{2\pi} \right)^{-1} \int_{-\infty}^{\infty} f_Y(\omega_j - y|\omega_{j-1}, \theta_Y) dy = \left( \epsilon \sqrt{2\pi} \right)^{-1}.
\]

Hence, the likelihood function is bounded above by \(A = \frac{1}{(\epsilon \sqrt{2\pi})^N} \prod_{j=1}^{N} S_j \left( \frac{\partial h}{\partial V_j} \right)^{-1} \), where \(N + 1\) is the number of observed stock data.
A.3 Proof of Proposition 3.3

According to the definition of the penalized likelihood function,

\[ L^S_p(\theta) = L^S(\theta)p_0(\sigma), \]

where \( p_0(\sigma) \) which is the inverted gamma density is a bounded function over \( \Theta \).

For \( \sigma > 0 \), the proof of Proposition 3.2 indicates that:

\[ L^S(\theta) \leq \frac{1}{(\sigma \sqrt{2\pi})^N} \prod_{j=1}^{N} S_j \left[ \frac{\partial h}{\partial V} \right]_{V=V_j, \theta}^{-1}. \]

Thus,

\[ L^S_p(\theta) \leq \frac{p_0(\sigma)}{(\sigma \sqrt{2\pi})^N} \prod_{j=1}^{N} S_j \left[ \frac{\partial h}{\partial V} \right]_{V=V_j, \theta}^{-1}. \]

Using the fact that \( p_0(\sigma) \sigma^{-N} \to 0 \) as \( \sigma \to 0^+ \), the proof is completed.

B The Jacobian of \( Q'_3 \)

This section presents the Jacobian of \( Q'_3 \), which is useful in the GEM algorithm for KJD structural models.

\[
\frac{\partial Q'_3}{\partial \mu} = \Delta t \sum_{t=1}^{N} \left\{ P_{1t} \left( \frac{(R_t - \mu_0)}{\sigma_0^2} \right) - P_{1t} \frac{f(w_t, m_1 + \eta_1 \sigma_0^2, \sigma_0^2)}{\Phi(d_{1t})} \right\},
\]

\[
+ (P_{11} \eta_1 - P_{20} \eta_2) + P_{2t} \frac{f(w_t, m_2 - \eta_2 \sigma_0^2, \sigma_0^2)}{\Phi((d_{2t})}.
\]

\[
\frac{\partial Q'_3}{\partial \sigma^2} = \frac{-b}{\sigma^2} + \frac{a \Delta t}{\sigma^4} + \frac{1}{2\sigma} \sum_{t=1}^{N} \left\{ P_{1t} \left( \frac{(R_t - \mu_0)(R_t - \mu_0 - \sigma_0^2)}{\sigma_0^2} \sqrt{\Delta t} - \frac{1}{\sigma} \right) \right\},
\]

\[
+ P_{1t} \left[ \frac{f(w_t, m_1 + \eta_1 \sigma_0^2/2, \sigma_0^2)(m_1 - w_t - 3\eta_1 \sigma_0^2/2 + \sigma_0^2)}{\sigma \Phi(d_{1t})} + \eta_1 \sigma \Delta t(\eta_1 - 1) \right],
\]

\[
+ P_{2t} \left[ \frac{f(w_t, m_2 - \eta_2 \sigma_0^2/2, \sigma_0^2)(w_t - m_2 - 3\eta_2 \sigma_0^2/2 + \sigma_0^2)}{\sigma \Phi(d_{2t})} + \eta_2 \sigma \Delta t(\eta_2 + 1) \right],
\]

\[
\frac{\partial Q'_3}{\partial \eta_1} = \sum_{t=1}^{N} P_{1t} \left[ \frac{1}{\eta_1} + m_1 - w_t + \frac{\eta_1 \sigma_0^2}{2} - \frac{\sigma_0^2 f(w_t, m_1 + \eta_1 \sigma_0^2/2, \sigma_0^2)}{\Phi(d_{1t})} \right],
\]

\[
\frac{\partial Q'_3}{\partial \eta_2} = \sum_{t=1}^{N} P_{2t} \left[ \frac{1}{\eta_2} + w_t - m_2 + \frac{\eta_2 \sigma_0^2}{2} - \frac{\sigma_0^2 f(w_t, m_2 - \eta_2 \sigma_0^2/2, \sigma_0^2)}{\Phi(d_{2t})} \right],
\]
where \( f(x, \mu, \sigma^2) \) is the density function of the normal distribution with mean \( \mu \) and variance \( \sigma^2 \), \( \mu_0 \) and \( \sigma_0 \) are defined in (28), and

\[
\begin{align*}
    d_{1t} &= \frac{w_t - m_1 - \eta_1 \sigma^2 \Delta t/2}{\sigma \sqrt{\Delta t}}, \\
    d_{2t} &= \frac{m_2 - w_t - \eta_2 \sigma^2 \Delta t/2}{\sigma \sqrt{\Delta t}}.
\end{align*}
\]

Then, we let the Jacobian \( J = \left[ \frac{\partial Q_3'}{\partial \mu} \frac{\partial Q_3'}{\partial \sigma^2} \frac{\partial Q_3'}{\partial m} \frac{\partial Q_3'}{\partial \eta_1} \frac{\partial Q_3'}{\partial \eta_2} \right]^T \). The Hessian matrix can then be approximated by a finite difference.

References


Table 7: MJD barrier-independent structural model

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of firms</th>
<th>$\mu$ (S.D.)</th>
<th>$\sigma$ (S.D.)</th>
<th>$\lambda$ (S.D.)</th>
<th>$k$ (S.D.)</th>
<th>$s$ (S.D.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1986</td>
<td>46</td>
<td>0.1656</td>
<td>0.1816</td>
<td>14.9095</td>
<td>0.0148</td>
<td>0.0304</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.1755)</td>
<td>(0.0370)</td>
<td>(18.2017)</td>
<td>(0.1001)</td>
<td>(0.0097)</td>
</tr>
<tr>
<td>1987</td>
<td>58</td>
<td>0.0791</td>
<td>0.2493</td>
<td>13.8500</td>
<td>-0.0068</td>
<td>0.0709</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.2705)</td>
<td>(0.0405)</td>
<td>(8.4049)</td>
<td>(0.0239)</td>
<td>(0.0367)</td>
</tr>
<tr>
<td>1988</td>
<td>76</td>
<td>0.1075</td>
<td>0.1771</td>
<td>14.1012</td>
<td>0.0037</td>
<td>0.0487</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.2795)</td>
<td>(0.0452)</td>
<td>(12.6177)</td>
<td>(0.0315)</td>
<td>(0.0733)</td>
</tr>
<tr>
<td>1989</td>
<td>77</td>
<td>0.1879</td>
<td>0.1601</td>
<td>8.8022</td>
<td>-0.0081</td>
<td>0.0535</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.1507)</td>
<td>(0.0362)</td>
<td>(8.0319)</td>
<td>(0.0299)</td>
<td>(0.0875)</td>
</tr>
<tr>
<td>1990</td>
<td>80</td>
<td>0.0044</td>
<td>0.1839</td>
<td>13.2197</td>
<td>-0.0026</td>
<td>0.0330</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.1595)</td>
<td>(0.0471)</td>
<td>(13.5161)</td>
<td>(0.0166)</td>
<td>(0.0100)</td>
</tr>
<tr>
<td>1991</td>
<td>105</td>
<td>0.3093</td>
<td>0.1962</td>
<td>14.9832</td>
<td>0.0026</td>
<td>0.0311</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.2030)</td>
<td>(0.0517)</td>
<td>(17.0436)</td>
<td>(0.0544)</td>
<td>(0.0127)</td>
</tr>
<tr>
<td>1992</td>
<td>116</td>
<td>0.1169</td>
<td>0.1759</td>
<td>12.2239</td>
<td>0.0091</td>
<td>0.0322</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.1561)</td>
<td>(0.0352)</td>
<td>(14.4408)</td>
<td>(0.0256)</td>
<td>(0.0119)</td>
</tr>
<tr>
<td>1993</td>
<td>137</td>
<td>0.2300</td>
<td>0.1735</td>
<td>10.8448</td>
<td>0.0226</td>
<td>0.0505</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.9302)</td>
<td>(0.0560)</td>
<td>(13.8672)</td>
<td>(0.1431)</td>
<td>(0.0904)</td>
</tr>
<tr>
<td>1994</td>
<td>152</td>
<td>-0.0415</td>
<td>0.1737</td>
<td>7.1313</td>
<td>0.0125</td>
<td>0.0421</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.2668)</td>
<td>(0.0393)</td>
<td>(9.6213)</td>
<td>(0.0654)</td>
<td>(0.0390)</td>
</tr>
<tr>
<td>1995</td>
<td>192</td>
<td>0.2807</td>
<td>0.1640</td>
<td>7.5671</td>
<td>0.0140</td>
<td>0.0405</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.2404)</td>
<td>(0.0365)</td>
<td>(9.5903)</td>
<td>(0.0559)</td>
<td>(0.0288)</td>
</tr>
<tr>
<td>1996</td>
<td>197</td>
<td>0.1717</td>
<td>0.1796</td>
<td>13.1410</td>
<td>0.0251</td>
<td>0.0367</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.3442)</td>
<td>(0.0491)</td>
<td>(16.1357)</td>
<td>(0.1352)</td>
<td>(0.0306)</td>
</tr>
<tr>
<td>Pooled</td>
<td>1236</td>
<td>0.1507</td>
<td>0.1791</td>
<td>11.2449</td>
<td>0.0111</td>
<td>0.0415</td>
</tr>
</tbody>
</table>
Table 8: MJD process vs. Diffusion process

<table>
<thead>
<tr>
<th>Year</th>
<th>( \frac{1}{N} \sum_{i=1}^{N} \frac{V_{D}(t_i) - V_{J}(t_i)}{V_{J}(t_i)} ) * (S.D.)</th>
<th>( \frac{\sigma_{D} - \sigma_{J}}{\sigma_{J}} ) * (S.D.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1986</td>
<td>-0.2641%</td>
<td>17.6097%</td>
</tr>
<tr>
<td></td>
<td>(2.2140%)</td>
<td>(25.9857%)</td>
</tr>
<tr>
<td>1987</td>
<td>-0.5210%</td>
<td>42.1537%</td>
</tr>
<tr>
<td></td>
<td>(2.1270%)</td>
<td>(22.8559%)</td>
</tr>
<tr>
<td>1988</td>
<td>-1.6997%</td>
<td>27.7795%</td>
</tr>
<tr>
<td></td>
<td>(8.1217%)</td>
<td>(46.7506%)</td>
</tr>
<tr>
<td>1989</td>
<td>-0.1422%</td>
<td>24.3013%</td>
</tr>
<tr>
<td></td>
<td>(15.9396%)</td>
<td>(33.9212%)</td>
</tr>
<tr>
<td>1990</td>
<td>-1.7238%</td>
<td>14.9217%</td>
</tr>
<tr>
<td></td>
<td>(9.0835%)</td>
<td>(23.4357%)</td>
</tr>
<tr>
<td>1991</td>
<td>0.1610%</td>
<td>14.2228%</td>
</tr>
<tr>
<td></td>
<td>(1.1942%)</td>
<td>(22.6125%)</td>
</tr>
<tr>
<td>1992</td>
<td>0.1258%</td>
<td>9.7964%</td>
</tr>
<tr>
<td></td>
<td>(0.6295%)</td>
<td>(12.0726%)</td>
</tr>
<tr>
<td>1993</td>
<td>-0.1930%</td>
<td>8.7326%</td>
</tr>
<tr>
<td></td>
<td>(2.6628%)</td>
<td>(15.3768%)</td>
</tr>
<tr>
<td>1994</td>
<td>0.0412%</td>
<td>11.7761%</td>
</tr>
<tr>
<td></td>
<td>(1.1801%)</td>
<td>(26.0321%)</td>
</tr>
<tr>
<td>1995</td>
<td>-0.2643%</td>
<td>15.9278%</td>
</tr>
<tr>
<td></td>
<td>(3.8455%)</td>
<td>(38.8957%)</td>
</tr>
<tr>
<td>1996</td>
<td>-0.1340%</td>
<td>19.8857%</td>
</tr>
<tr>
<td></td>
<td>(7.9959%)</td>
<td>(25.2412%)</td>
</tr>
<tr>
<td>Pooled</td>
<td>0.3110%</td>
<td>17.0088%</td>
</tr>
</tbody>
</table>

*where \( V_{D} \) and \( \sigma_{D} \) are respectively the firm value and volatility estimated from the diffusion model while \( V_{J} \) and \( \sigma_{J} \) are estimated from the MJD model. \( N \) is the number of time points.