Parametric Regularized Calibration of Merton Jump-Diffusion Model with Relative Entropy: What Difference Does It Make?

Kazuhisa Matsuda

Department of Economics
The Graduate Center, The City University of New York,
365 Fifth Avenue, New York, NY 10016-4309
Email: maxmatsuda@hotmail.com
http://www.maxmatsuda.com/

July 18 2005

Abstract

This paper answers to the simple question, "What difference does the regularization with relative entropy make?" For this purpose, we parametrically calibrate the Merton (1976) jump-diffusion model the S&P 500 futures options with or without regularization and judge the impact of the regularization. The calibration result indicates that with or without regularization, calibrated risk-neutral parameters, calibrated log return probability densities, and calibrated Lévy measures are not significantly different using the risk-neutral prior except for near maturity options. Another finding is that the regularized calibration with the statistical prior is not implementable because of the remarkable deviation of the statistical prior from the risk-neutral prior.

[1] Introduction

A calibration problem is an inverse problem which tries to identify (i.e. back out) a vector of parameters $\theta^{\mathbb{Q}}$ which produce model option prices consistent with market (i.e. observed) option prices. When the objective function which is usually the sum of squared dollar pricing errors between market prices and model prices is not convex, the calibration problem is an illposed problem. In this illposed calibration problem, some solution can always be found (i.e. non-uniqueness), but the solution obtained is very sensitive to the initial values (i.e. instability) and it is not the global minimum with very high likelihood. Traditionally in the field of finance, the gradient descent algorithms such as the BFGS method are generally used to solve this illposed problem. To raise a few examples, Bates (1996), Bakshi, Cao, Chen (1997), Dumas, Fleming, and Whaley (1998), Carr, Chang, and Madan (1998b), and Carr, Geman, Madan, and Yor (2002).

To overcome above mentioned difficulties in illposed problems, statisticians and mathematicians have long been using regularization methods. Regularization methods are not the methods to locate the global solution, but they are the methods to enhance the uniqueness and the stability of the calibration solution by sacrificing its precision. Cont and Tankov (2004a, b) choose the regularization with the relative entropy $\mathcal{E}(\mathbb{Q}|\mathbb{P})$ which is a measure of distance between two probability measures and describes the amount of inefficiency to assume that the true distribution is $\mathbb{Q}$ when the true distribution of the random variable X is $\mathbb{P}$. It is very convenient that the relative entropy for Lévy processes can be explicitly expressed in terms of their Lévy measures. It is important to realize that using the relative entropy and using the prior mean the introduction of the bias of the calibration solution toward the prior to gain the numerical stability and the uniqueness by making the objective function more convex. It implies that the user has some or strong belief in the use of the prior (i.e. otherwise, why do you bother?). Cont and Tankov develop the non-parametric regularized calibration method for Merton (1976) jump-diffusion model with the relative entropy.

The goal of this paper is to give an answer to the following simple question, “What difference does the regularization with relative entropy make?” For this purpose, we calibrate the Merton (1976) jump-diffusion model which is a Lévy model with occasional but rare jumps to the S&P 500 futures options with or without regularization. The impact of the regularization over the unregularized calibration is judged by the calibrated log return probability density and the calibrated Lévy measure.

Note that we apply Cont and Tankov’s (2004 a, b) regularization method with the relative entropy to the index options *parametrically* with the MJD model. This is different from Cont and Tankov’s (2004 a, b) research which is to calibrate Merton jump-diffusion model non-parametrically with the relative entropy regularization.

The calibration result suggests that with or without regularization, calibrated risk-neutral parameters, calibrated log return probability densities, and calibrated Lévy measures are not significantly different. It seems that Lévy measures are more sensitive to the regularization than log return probability measures. Notice also that the difference in calibrated parameters between the regularized and the unregularized become more pronounced especially for near maturity options.

This paper is organized as follows. Section 2 gives the detailed description of the Merton jump-diffusion model. Section 3 presents the (unregularized) calibration problem as an inverse problem and as an illposed problem due to the non-convexity of the objective function. Section 4 briefly reviews Cont and Tankov’s (2004a, b) method of the regularized calibration with the relative entropy which tries to achieve a unique solution and a stable solution. Section 5 describes the S&P 500 futures option data set and obtains two different prior probability measures. One is the statistical prior and the other is the risk-neutral prior. Section 6 provides our main empirical result of the difference between the regularized calibration and the unregularized calibration in terms of the calibrated parameter vector, the calibrated log return probability density, and the calibrated Lévy measure. Section 7 concludes.

[2] Merton Jump Diffusion (MJD) Model (1976)

Consider a fixed filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty]}, \mathbb{P})$. A jump diffusion process $(X_{t\in[0,\infty]})$ with the Lévy triplet $(A_X = \sigma^2, \ell_X = \lambda f(x), \gamma_X = 0)$ is defined as a Brownian motion plus a compound Poisson process:

$$(X_{t\in[0,\infty]}) \equiv (\sigma B_{t\in[0,\infty]}) + \sum_{i=1}^{N_t} X_i \, , \tag{1}$$

where $(\sigma B_{t\in[0,\infty]})$ is a multiplicative Brownian motion with the Lévy triplet $(A_B = \sigma^2, \ell_B = 0, \gamma_B = 0)$, $\sum_{i=1}^{N_t} X_i$ is a compound Poisson process with the Lévy triplet $(A_C = 0, \ell_C = \lambda f(x), \gamma_C = 0)$ which is the sum of $i.i.d.$ jumps X_i from the jump size probability density $f(x)$, and $(N_{t\in[0,\infty]})$ is a Poisson process with the intensity $\lambda \in \mathbb{R}^+$ which counts the number of random arrival times T_k of an event in the time interval $[0,t]$:

$$N_t = \sum_{k\geq 1} 1_{t\geq T_k} \, . \tag{2}$$

Note that a Poisson process $(N_{t\in[0,\infty]})$ and the jumps sizes $(X_i)_{i\geq 1}$ are assumed to be independent. This jump diffusion process $(X_{t\in[0,\infty]})$ possesses the following properties. It is a jump Lévy process, but not a pure jump Lévy process because the Gaussian variance term of the jump diffusion process A_X is non-zero. In other words, the process contains a Brownian motion[1]. The Lévy measure of the jump diffusion process is given by:

$$\ell_X(x;\lambda) = \lambda f(x) \, , \tag{3}$$

where $f(x)$ is the jump size probability density. The total mass of the Lévy measure of the jump diffusion process is the intensity parameter λ because a Lévy measure $\ell(x)$ measures the arrival rate of jumps:

$$\int_{-\infty}^{\infty} \ell_X(x)dx = \int_{-\infty}^{\infty} \lambda f(x)dx = \lambda \int_{-\infty}^{\infty} f(x)dx = \lambda < \infty \, ,$$

which is finite because the number of arrivals of an event is almost surely finite for any $t > 0$ including an infinite time horizon $t = \infty$[2]. In other words, the jump diffusion process is a finite activity Lévy process which means that the process has finite number of small jumps and finite number of large jumps. The jump diffusion process is also a Lévy process of infinite variation in the interval $[0, \infty)$ because $A_X = \sigma^2 \neq 0$.[3]

---

[1] Consult section 3.7.1 of Matsuda (2005a).
[2] The number (6) of theorem 4.10 of Matsuda (2005a).
[3] Consult theorem 3.11 of Matsuda (2005a).

Merton jump diffusion (MJD) model specifies the log return jump size density as the normal, i.e. $X_i \sim i.i.d.\ Normal(\mu, \delta^2)$:

$$f_{MJD}(x) = \frac{1}{\sqrt{2\pi\delta^2}} \exp\left\{ -\frac{\left(x-\mu\right)^2}{2\delta^2} \right\}. \tag{4}$$

Thus, the Lévy measure in MJD model can be expressed as:

$$\ell_{MJD,X}(x; \lambda, \mu, \delta) = \frac{\lambda}{\sqrt{2\pi\delta^2}} \exp\left\{ -\frac{\left(x-\mu\right)^2}{2\delta^2} \right\}. \tag{5}$$

The characteristic function of MJD process can be obtained by the use of the Lévy-Khinchin representation[4] as:

$$\phi_{MJD,X}(\omega; \sigma, \lambda, \mu, \delta) = \exp\left[t \left\{ -\frac{\sigma^2\omega^2}{2} + \lambda\left(\phi_f(\omega) - 1\right) \right\} \right], \tag{6}$$

where ϕ_f is the characteristic function of the jump size density:

$$\phi_f(\omega) = \exp\left(i\omega\mu - \frac{\delta^2\omega^2}{2} \right).$$

The probability density of the MJD process can be computed using the conditionally normal property of the jump diffusion process of the equation (1)[5]:

$$\mathbb{P}_{MJD}(x_t; \sigma, \lambda, \mu, \delta) = \sum_{j=0}^{\infty} \mathbb{P}(x_t \left| N_t = j \right.) \mathbb{P}_{Poisson}(N_t = j)$$

$$\mathbb{P}_{MJD}(x_t; \sigma, \lambda, \mu, \delta) = \sum_{j=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^j}{j!} \frac{1}{\sqrt{2\pi\left(\sigma^2 t + j\delta^2\right)}} \exp\left\{ -\frac{\left(x_t - j\mu\right)^2}{2\left(\sigma^2 t + j\delta^2\right)} \right\}. \tag{7}$$

Its standardized moments are computed by:

$$E[X_t] = \lambda t \mu, \tag{8}$$

$$Variance[X_t] = (\sigma^2 + \lambda\delta^2 + \lambda\mu^2)t,$$

---

[4] Consult the section 4.3.4 of Matsuda (2005a) where the characteristic function of a compound Poisson process is obtained by using the Lévy-Khinchin representation for the finite variation processes.

[5] This computation involves the series expansion rather than the integration because a compound Poisson process is a continuous time stochastic process with the discontinuous sample paths.

$$Skewness[X_t] = \frac{t\lambda\mu(\mu^2 + 3\delta^2)}{Variance[X_t]^{3/2}},$$

$$Excess\ Kurtosis[X_t] = \frac{t\lambda(\mu^4 + 3\delta^4 + 6\mu^2\delta^2)}{Variance[X_t]^2}.$$

These standardized moments indicate that μ is a skewness parameter with $\mu = 0$ producing the symmetric probability density. Larger values for λ and σ lead to the larger variance and smaller excess kurtosis of the probability density.

MJD model specifies the asset price dynamics $(S_{t\in[0,T]})$ defined on a filtered risk neutral probability space $(\Omega, \mathcal{F}_{t\in[0,T]}, \mathbb{Q})$ as an exponential of a Lévy process $(L_{t\in[0,T]})$:

$$S_t = S_0 \exp(L_t),$$

where the choice of the Lévy process is the jump diffusion process plus the drift $r - \varpi_{MJD,\mathbb{Q}}$:

$$L_t \equiv (r - \varpi_{MJD,\mathbb{Q}})t + MJD(x_t; \sigma_{\mathbb{Q}}, \lambda_{\mathbb{Q}}, \mu_{\mathbb{Q}}, \delta_{\mathbb{Q}}), \tag{9}$$

where $r \in \mathbb{R}^+$ is the instantaneous risk-free interest rate and all parameters are under the risk neutral probability measure $\mathbb{Q}$. The term $\varpi_{MJD,\mathbb{Q}}$ is the convexity correction which takes the following form in the MJD model:

$$\varpi_{MJD,\mathbb{Q}} = \lambda_{\mathbb{Q}} \left\{ \exp\left(\mu_{\mathbb{Q}} + \frac{\delta_{\mathbb{Q}}{}^2}{2} \right) - 1 \right\} + \frac{\sigma_{\mathbb{Q}}{}^2}{2}. \tag{10}$$

Defining the log return (i.e. log price relative) of the asset price as $R_t \equiv \ln(S_t / S_0)$ and using the equation (9):

$$x_t = R_t - (r - \varpi_{MJD,\mathbb{Q}})t. \tag{11}$$

Since obviously the drift $r - \varpi_{MJD,\mathbb{Q}}$ is deterministic, the probability density of the log return in the MJD model under the risk neutral probability measure $\mathbb{Q}$ can be expressed using the probability density (7) and the relationship (11) as:

$$\mathbb{Q}_{MJD}(R_t; \sigma, \lambda, \mu, \delta) = \sum_{j=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^j}{j!} \frac{1}{\sqrt{2\pi\left(\sigma^2 t + j\delta^2\right)}} \exp\left\{ -\frac{\left(x_t - j\mu\right)^2}{2\left(\sigma^2 t + j\delta^2\right)} \right\}, \tag{12}$$

with the equation (11). Note that all parameters in the density (12) are under the risk neutral probability measure $\mathbb{Q}$.

[3] Calibration without Regularization

[3.1] Calibration: An Inverse Problem

Calibration is an inverse problem. Consider two problems. One of which is named as a direct problem and the other is named as an inverse problem. In the context of option pricing, a direct problem is formulated as:

$$C_i^{\text{model}}(\theta^{\mathbb{Q}}, S_t, K_i, T-t) = e^{-r(T-t)} E^{\mathbb{Q}}\left[\left(S_T - K_i\right)^+ \middle| \mathcal{F}_t\right], \tag{13}$$

where European call option prices C_i^{model} across different strikes $K_{i\in I}$ are calculated given a model, a vector of risk-neutral model parameters $\theta^{\mathbb{Q}}$, and variables such as a spot asset price S_t and a maturity $T-t$. Following the martingale asset pricing, a European call price is equal to the discounted value of the terminal payoff under a risk-neutral probability measure $\mathbb{Q}$. For example, $\theta^{\mathbb{Q}} = (\sigma^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \mu^{\mathbb{Q}}, \delta^{\mathbb{Q}})$ in the MJD case. An inverse problem (called a parameter identification problem) is formulated as the reverse of this procedure. It is to identify (i.e. back out) a vector of parameters $\theta^{\mathbb{Q}}$ which produce model option prices consistent with market (i.e. observed) option prices:

$$C_i^{\text{model}}(\theta^{\mathbb{Q}}) = C_i^{\text{market}}. \tag{14}$$

The exact match of the model prices to the market prices is not necessary because of the noise (i.e. bid-ask spreads) contained in the market option prices $(C_{i\in I}^{\text{market}})$. Thus, the calibration problem becomes a best approximation problem between market prices and model prices which is done using a nonlinear least squares (NLS):

$$\theta^{\mathbb{Q}} = \arg\min_{\theta^{\mathbb{Q}}} \sum_{i=1}^{N} \left| C_i^{\text{model}}(\theta^{\mathbb{Q}}) - C_i^{\text{market}} \right|^2, \tag{15}$$

where the risk-neutral parameter vector $\theta^{\mathbb{Q}}$ is chosen by minimizing the sum of squared dollar pricing errors between market prices and model prices. The gradient descent algorithms such as the BFGS method are generally used to solve the optimization problem of the equation (15).

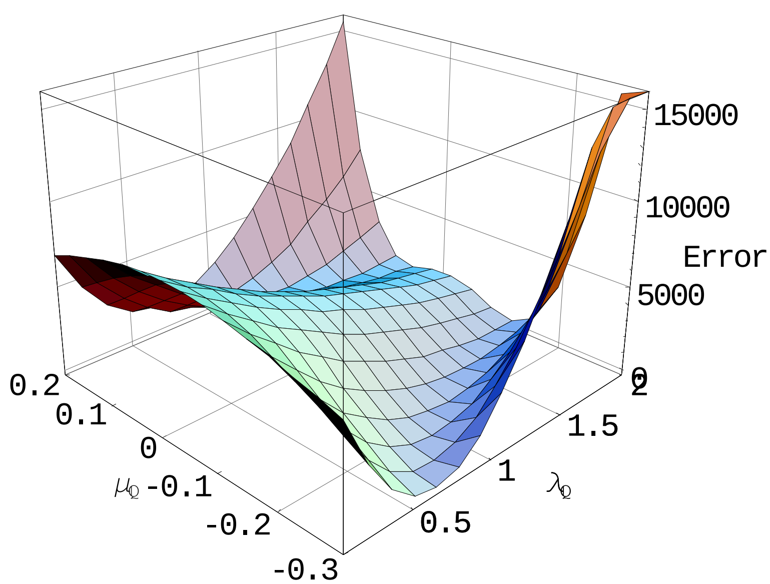
Followings are examples of literatures which use this NLS without regularization. Bates (1996) calibrates stochastic volatility/jump models to a currency option, Bakshi, Cao, Chen (1997) calibrates stochastic volatility/jump models to an index option,and Dumas, Fleming, and Whaley (1998) uses unregularized calibration for fitting a deterministic volatility function. Carr, Chang, and Madan (1998b) calibrate the Variance Gamma model to an index option through the maximum likelihood estimation which is equivalent to the NLS. Carr, Geman, Madan, and Yor (2002) calibrates the extended CGMY model to individual stock options and index options.

[3.2] Calibration: An Illposed Problem

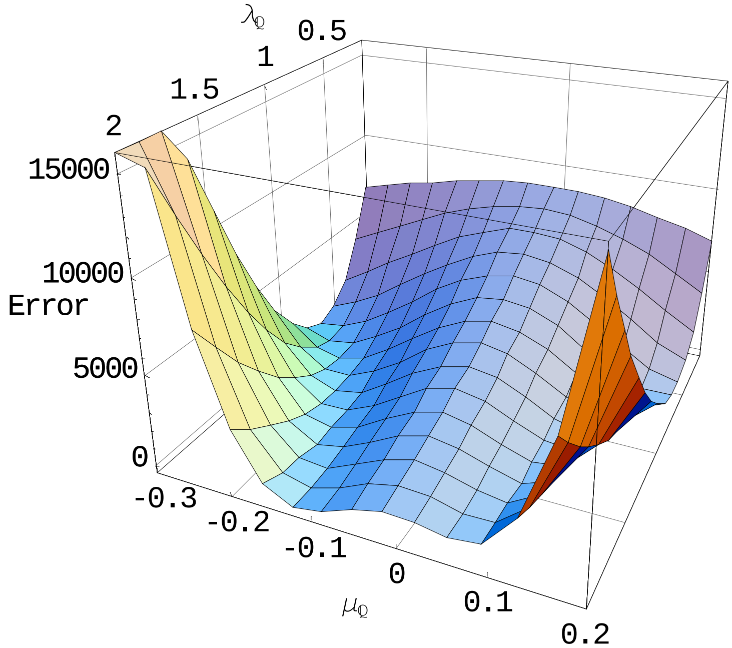
According to Cont and Tankov (2004a, b), an illposed problem is a problem which possesses the following properties. Firstly, an illposed problem may not have a solution or may have an infinite number of solutions. Secondly, when a solution or solutions of an illposed problem exists (exist) and if some type of an additional criterion is used to choose a solution, it is very sensitive to the initial values. Thirdly, when a solution of an illposed problem exists, it might be difficult to obtain it because it is likely to get stuck at local minima (due to the non-convex objective function).

Our interest of the calibration problem of the equation (15) is an illposed problem whose illposedness is solely caused by the non-convex objective function. One example of the sum of squared dollar pricing error function for the MJD model is illustrated in Figure 1 as a function of parameters $\lambda^{\mathbb{Q}}$ and $\mu^{\mathbb{Q}}$ with other parameters being fixed. Due to its non-convex nature, the optimization problem (15) possesses the following illposedness. Firstly, some solution can always be found (this is not necessarily a good thing). But, secondly, the solution obtained is very sensitive to the initial values. In other words, the solution is very instable. Thirdly, it is highly likely that the solution obtained is not the global minimum.

Note that the calibration problem of the equation (15) is not the only illposed problem. Illposed problems are everywhere which include the numerous maximum likelihood estimation problems. In the past, researchers have used an ad hoc treatment for illposed problems such as repeating the optimization procedure with various initial values.



A) From one angle.



B) From another angle.

Figure 1 3D plot of the MJD error function $\sum_{i=1}^{N}\left|C_i^{\text{model}}(\theta^{\mathbb{Q}})-C_i^{\text{market}}\right|^2$ The sum of squared dollar pricing error function is plotted as a function of parameters $\lambda^{\mathbb{Q}}$ and $\mu^{\mathbb{Q}}$ with other parameters being fixed.

[4] Regularized Calibration

[4.1] Relative Entropy

We saw in the previous section that the NLS calibration problem has the difficulty in achieving a unique solution and a stable solution. To overcome this issue, regularization methods have been developed. Engl, Hanke, and Neubauer (1996) give a brief summary of regularization methods. In this paper, we focus on the regularization with relative entropy which is used by Cont and Tankov (2004a, b). Note that Cont and Tankov are the first to use relative entropy regularization for the calibration of the exponential Lévy models, but they are not the first in using relative entropy regularization in the finance context. We believe it was Avellaneda, Friedman, Holmes, and Samperi (1997) who used the relative entropy regularization for the calibration of volatility surfaces.

The relative entropy is a measure of distance between two probability measures which is expressed as the expected value of the logarithm of the likelihood ratio. If we knew that the true distribution of the random variable X is $\mathbb{P}$, the relative entropy $\mathcal{E}(\mathbb{Q}|\mathbb{P})$ describes the amount of inefficiency to assume that the true distribution is $\mathbb{Q}$.

DEFINITION: Let $(X_{T\in[0,\infty]})$ be a real-valued rcll process defined on a filtered probability space $(\Omega, \mathcal{F}_{T\in[0,\infty]}, \mathbb{P})$. Let $\mathbb{P}$ and $\mathbb{Q}$ be two equivalent probability measures on $(\Omega, \mathcal{F}_{T\in[0,\infty]})$. The relative entropy or Kullback-Leibler distance between two probability measures $\mathbb{P}$ and $\mathbb{Q}$ is defined as:

$$\mathcal{E}(\mathbb{Q}|\mathbb{P}) = E^{\mathbb{Q}}[\ln \frac{d\mathbb{Q}}{d\mathbb{P}}] = E^{\mathbb{P}}[\frac{d\mathbb{Q}}{d\mathbb{P}} \ln \frac{d\mathbb{Q}}{d\mathbb{P}}]. \tag{16}$$

Relative entropy is a convex function of $\mathbb{Q}$ because after a rearrangement:

$$\mathcal{E}(\mathbb{Q}|\mathbb{P}) = E^{\mathbb{P}}[f(\frac{d\mathbb{Q}}{d\mathbb{P}})], \tag{17}$$

where $f(x) = x \ln x$ is a strictly convex function as illustrated in Figure 2. This strict convexity of $\mathcal{E}(\mathbb{Q}|\mathbb{P})$ plays a crucial role in enhancing the uniqueness of a solution of an illposed problem. Relative entropy is always non-negative (i.e. $\mathcal{E}(\mathbb{Q}|\mathbb{P}) \geq 0$) and $\mathcal{E}(\mathbb{Q}|\mathbb{P}) = 0$ if and only if $d\mathbb{Q}/d\mathbb{P} = 1$ almost surely. Cover and Thomas (1991) provide the proof.

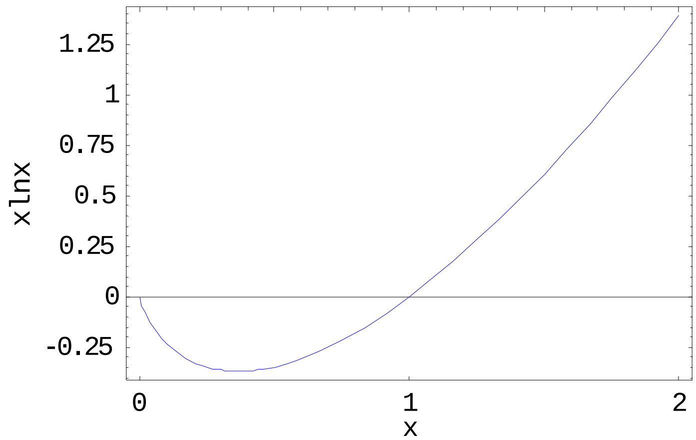


Figure 2 Strictly convex function $f(x) = x \ln x$

[4.2] Relative Entropy for Lévy Processes

Very conveniently, relative entropy for Lévy processes can be explicitly expressed in terms of their Lévy measures. The following useful theorem corresponds to the proposition 2 of Cont and Tankov (2004b).

THEOREM: Relative Entropy for Lévy Processes

Let $(X_{T\in[0,\infty]})$ be a real-valued Lévy process defined on a filtered probability space $(\Omega, \mathcal{F}_{T\in[0,\infty]}, \mathbb{P})$ with the Lévy triplet $(A^{\mathbb{P}}, \ell^{\mathbb{P}}, \gamma^{\mathbb{P}})$. Define a probability measure $\mathbb{Q} \sim \mathbb{P}$ under which a Lévy process $(X_{T\in[0,\infty]})$ is described by the triplet $(A^{\mathbb{Q}}, \ell^{\mathbb{Q}}, \gamma^{\mathbb{Q}})$. Note that $A = A^{\mathbb{P}} = A^{\mathbb{Q}}$ because $\mathbb{Q} \sim \mathbb{P}$. Then, for every time horizon $T \in [0,\infty]$, the relative entropy of $\mathbb{Q}|\mathcal{F}_T$ with respect to $\mathbb{P}|\mathcal{F}_T$ can be expressed as:

$$\begin{aligned}\mathcal{E}_T(\mathbb{Q}|\mathbb{P}) = \frac{T}{2A}\left\{\gamma^{\mathbb{Q}} - \gamma^{\mathbb{P}} - \int_{-1}^{1} x(\ell^{\mathbb{Q}} - \ell^{\mathbb{P}})(dx)\right\}^2 1_{A\neq 0} \\ + T\int_{-\infty}^{\infty}\left(\frac{d\ell^{\mathbb{Q}}}{d\ell^{\mathbb{P}}}\ln\frac{d\ell^{\mathbb{Q}}}{d\ell^{\mathbb{P}}} + 1 - \frac{d\ell^{\mathbb{Q}}}{d\ell^{\mathbb{P}}}\right)\ell^{\mathbb{P}}(dx). \qquad (18)\end{aligned}$$

When $\mathbb{P}$ and $\mathbb{Q}$ are both the risk-neutral martingale measures, for every time horizon $T \in [0,\infty]$, the relative entropy of $\mathbb{Q}|\mathcal{F}_T$ with respect to $\mathbb{P}|\mathcal{F}_T$ can be reduced to:

$$\begin{aligned}\mathcal{E}_T(\mathbb{Q}|\mathbb{P}) = \frac{T}{2A}\left\{\int_{-\infty}^{\infty} (e^x - 1)(\ell^{\mathbb{Q}} - \ell^{\mathbb{P}})(dx)\right\}^2 1_{A\neq 0} \\ + T\int_{-\infty}^{\infty}\left(\frac{d\ell^{\mathbb{Q}}}{d\ell^{\mathbb{P}}}\ln\frac{d\ell^{\mathbb{Q}}}{d\ell^{\mathbb{P}}} + 1 - \frac{d\ell^{\mathbb{Q}}}{d\ell^{\mathbb{P}}}\right)\ell^{\mathbb{P}}(dx). \qquad (19)\end{aligned}$$

Example of Relative Entropy: Brownian Motion with Drift Let $(X_{T\in[0,\infty]})$ be a Brownian motion with drift defined on a filtered probability space $(\Omega, \mathcal{F}_{T\in[0,\infty]}, \mathbb{P})$ with the Lévy triplet $(A^{\mathbb{P}} = (\sigma^{\mathbb{P}})^2, \ell^{\mathbb{P}} = 0, \gamma^{\mathbb{P}} = \mu^{\mathbb{P}})$:

$$X_T{}^{\mathbb{P}} = \mu^{\mathbb{P}} T + \sigma^{\mathbb{P}} B_T{}^{\mathbb{P}} .$$

Define a probability measure $\mathbb{Q} \sim \mathbb{P}$ under which a Brownian motion with drift $(X_{T\in[0,\infty]})$ is described by the triplet $(A^{\mathbb{Q}} = (\sigma^{\mathbb{Q}})^2, \ell^{\mathbb{Q}} = 0, \gamma^{\mathbb{Q}} = \mu^{\mathbb{Q}})$. Note that $A^{\mathbb{P}} = A^{\mathbb{Q}} = \sigma^2$ because $\mathbb{Q} \sim \mathbb{P}$. Then, for every time horizon $T \in [0,\infty]$, the relative entropy of $\mathbb{Q}|\mathcal{F}_T$ with respect to $\mathbb{P}|\mathcal{F}_T$ can be expressed as by applying the equation (18):

$$\mathcal{E}_T(\mathbb{Q}|\mathbb{P}) = \frac{T}{2\sigma^2}\left\{\mu^{\mathbb{Q}} - \mu^{\mathbb{P}}\right\}^2 = \frac{1}{2}\left(\frac{\mu^{\mathbb{Q}} - \mu^{\mathbb{P}}}{\sigma}\right)^2 T . \qquad (20)$$

Thus, in Gaussian case, the relative entropy function is symmetric:

$$\mathcal{E}_T(\mathbb{Q}|\mathbb{P}) = \mathcal{E}_T(\mathbb{P}|\mathbb{Q}),$$

because $(\mu^{\mathbb{Q}} - \mu^{\mathbb{P}})^2 = (\mu^{\mathbb{P}} - \mu^{\mathbb{Q}})^2$. The relative entropy function of the equation (20) is plotted in Figure 3 by setting $T = 1$ and $\sigma = 1$ for simplicity. Notice its strict convexity.

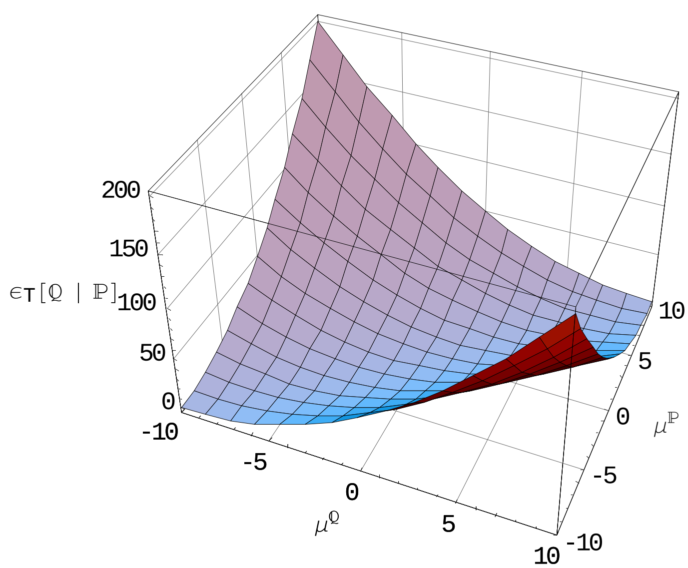


Figure 3 Relative entropy function for Gaussian case We set $T = 1$ and $\sigma = 1$ for simplicity.

Example of Relative Entropy: Merton jump-diffusion process Let $(X_{T\in[0,\infty]})$ be a jump diffusion process of the equation (1) defined on a filtered probability space $(\Omega, \mathcal{F}_{T\in[0,\infty]}, \mathbb{P})$ with the Lévy triplet $(A^{\mathbb{P}} = (\sigma^{\mathbb{P}})^2, \ell^{\mathbb{P}} = \lambda^{\mathbb{P}} f(x)^{\mathbb{P}}, \gamma^{\mathbb{P}} = 0)$. Define a risk-neutral martingale probability measure $\mathbb{Q} \sim \mathbb{P}$ under which a jump diffusion process $(X_{T\in[0,\infty]})$ is described by the triplet $(A^{\mathbb{Q}} = (\sigma^{\mathbb{Q}})^2, \ell^{\mathbb{Q}} = \lambda^{\mathbb{Q}} f(x)^{\mathbb{Q}}, \gamma^{\mathbb{Q}} = 0)$. Note that $A^{\mathbb{P}} = A^{\mathbb{Q}} = \sigma^2$ because $\mathbb{Q} \sim \mathbb{P}$. From the equation (5), the Lévy measures take the form:

$$\ell^{\mathbb{P}} = \frac{\lambda^{\mathbb{P}}}{\sqrt{2\pi(\delta^{\mathbb{P}})^2}} \exp\left\{ -\frac{\left(x-\mu^{\mathbb{P}}\right)^2}{2(\delta^{\mathbb{P}})^2} \right\}, \tag{21}$$

$$\ell^{\mathbb{Q}} = \frac{\lambda^{\mathbb{Q}}}{\sqrt{2\pi(\delta^{\mathbb{Q}})^2}} \exp\left\{ -\frac{\left(x-\mu^{\mathbb{Q}}\right)^2}{2(\delta^{\mathbb{Q}})^2} \right\}. \tag{22}$$

Then, after a little bit of algebra, the relative entropy of $\mathbb{Q}|\mathcal{F}_T$ with respect to $\mathbb{P}|\mathcal{F}_T$ can be expressed as by applying the equation (19) with the Lévy measures (21) and (22):

$$\begin{aligned} T^{-1}\mathcal{E}(\mathbb{Q}|\mathbb{P}) = \frac{1}{2\sigma^2}&\left\{ \lambda^{\mathbb{Q}}\left(\mathrm{e}^{\mu^{\mathbb{Q}}+\frac{\delta^{\mathbb{Q}2}}{2}} - 1 \right) - \lambda^{\mathbb{P}}\left(\mathrm{e}^{\mu^{\mathbb{P}}+\frac{\delta^{\mathbb{P}2}}{2}} - 1 \right) \right\}^2 \\ &+ \lambda^{\mathbb{Q}} \ln\left(\frac{\lambda^{\mathbb{Q}}\delta^{\mathbb{P}}}{\lambda^{\mathbb{P}}\delta^{\mathbb{Q}}} \right) + \lambda^{\mathbb{P}} + \lambda^{\mathbb{Q}}\left(-\frac{3}{2} + \frac{(\mu^{\mathbb{Q}}-\mu^{\mathbb{P}})^2 + \delta^{\mathbb{Q}2}}{2\delta^{\mathbb{P}2}} \right). \end{aligned} \tag{23}$$

The relative entropy function of the equation (23) is plotted in Figure 4 by setting $T = 0.25$, $\sigma = 0.1$, $\delta^{\mathbb{Q}} = 0.15$, $\lambda^{\mathbb{P}} = 1$, $\mu^{\mathbb{P}} = -0.1$, and $\delta^{\mathbb{P}} = 0.1$. Cont and Tankov (2004b) points out that the relative entropy function in the MJD case of the equation (23) is not a concave function in $\lambda^{\mathbb{Q}}$ $\mu^{\mathbb{Q}}$ $\delta^{\mathbb{Q}}$ because its Lévy measure is a nonlinear function in λ, μ, δ.

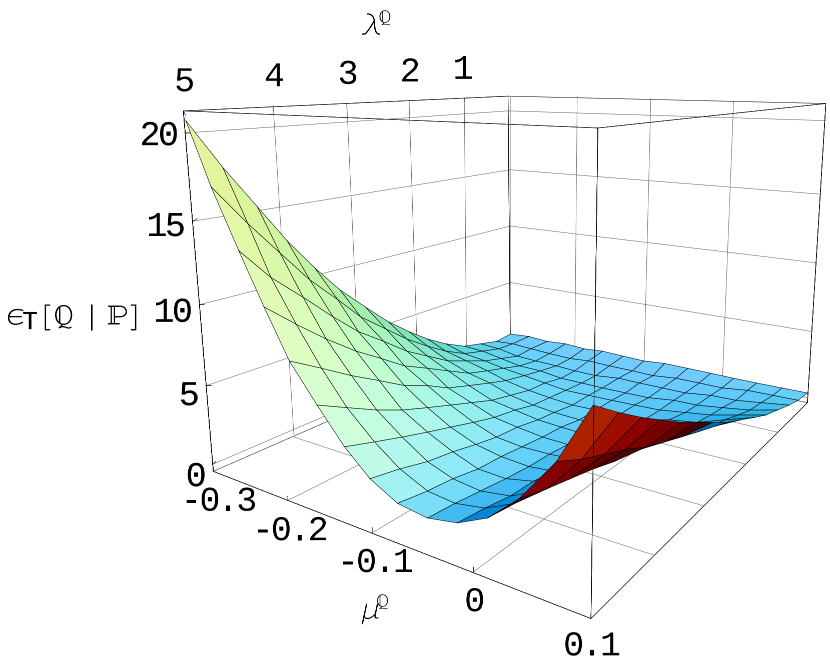


Figure 4 Relative entropy function for Merton Jump-Diffusion Process We set $T = 0.25$, $\sigma = 0.1$, $\delta^{\mathbb{Q}} = 0.15$, $\lambda^{\mathbb{P}} = 1$, $\mu^{\mathbb{P}} = -0.1$, and $\delta^{\mathbb{P}} = 0.1$.

[4.3] Regularized Calibration

The uniqueness and the stability of the solution of the NLS calibration problem of the equation (15) can be augmented by adding the relative entropy term:

$$\theta^{\mathbb{Q}} = \arg\min_{\theta^{\mathbb{Q}}} \sum_{i=1}^{N} \left| C_i^{\text{model}}(\theta^{\mathbb{Q}}) - C_i^{\text{market}} \right|^2 + \alpha \mathcal{E}_T(\mathbb{Q} | \mathbb{P}), \tag{24}$$

where α is a regularization parameter. The convexity of the relative entropy $\mathcal{E}_T(\mathbb{Q}|\mathbb{P})$ makes the non-convex objective function more convex, thus, enhancing the uniqueness and the stability of the solution. According to Cont and Tankov (2004a, b), the relative entropy $\mathcal{E}_T(\mathbb{Q}|\mathbb{P})$ remains convex in the neighborhood of its global minimum as long as the parameterization is well-behaved when $\mathcal{E}_T(\mathbb{Q}|\mathbb{P})$ is not strictly convex with respect to $\theta^{\mathbb{Q}}$.

By adding and minimizing the relative entropy $\mathcal{E}_T(\mathbb{Q}|\mathbb{P})$, we are making the risk-neutral martingale measure $\mathbb{Q}$ as close as possible to the prior (i.e. statistical) measure $\mathbb{P}$. For example, consider a case where $\mathbb{Q}_1$ exactly matches the market option prices. but $\mathbb{Q}_1$ is far away from the prior $\mathbb{P}$:

$$\sum_{i=1}^{N} \left| C_i^{\text{model}}(\theta^{\mathbb{Q}_1}) - C_i^{\text{market}} \right|^2 + \alpha \mathcal{E}_T(\mathbb{Q}_1 | \mathbb{P}) = \alpha \mathcal{E}_T(\mathbb{Q}_1 | \mathbb{P}),$$

In this case, we are willing to sacrifice the precision of the calibration and choose $\mathbb{Q}_2$ which is closer to the prior $\mathbb{P}$ within the error bounds of the bid-ask spread:

$$\sum_{i=1}^{N} \left| C_i^{\text{model}}(\theta^{\mathbb{Q}_2}) - C_i^{\text{market}} \right|^2 + \alpha \mathcal{E}_T(\mathbb{Q}_2 \| \mathbb{P}) < \alpha \mathcal{E}_T(\mathbb{Q}_1 \| \mathbb{P}) .$$

In other words, the regularization by the relative entropy means the introduction of the bias of the calibrated parameter vector $\theta_{\mathbb{Q}}$ under the risk-neutral martingale measure $\mathbb{Q}$ toward the prior parameter vector $\theta_{\mathbb{P}}$ under the statistical measure $\mathbb{P}$ (i.e. time series behavior of underlying prices) rather than relying solely on the new information contained in the quoted option prices.

There are two important parameters which should be chosen with care. Those are the prior parameter vector $\theta_{\mathbb{P}}$ and the regularization parameter α. With respect to the choice of the prior $\theta_{\mathbb{P}}$, Cont and Tankov (2004a, b) suggest the followings. The first is to employ the historical prior which is estimated using the time series of the underlying price by the statistical method (i.e. maximum likelihood estimation). The second approach which does not require the time series data is to run an unregularized calibration of the equation (15) and use the successively updated unregularized risk-neutral parameter solution $\theta_{\mathbb{Q}}$ as the prior $\theta_{\mathbb{P}}$. The third is to use the long-run average of the unregularized risk-neutral parameter solutions $\theta_{\mathbb{Q}}$. The third approach is preferable over the second because the role of the prior is to enforce the stability in the solution.

A regularization parameter α is a weight assigned to the relative entropy $\mathcal{E}_T(\mathbb{Q} \| \mathbb{P})$ and cannot be approximated by *a priori* fixed number because of its dependence on the level of noise present in the data. When α is large, we relatively trust the prior information more than the new information contained in the market option prices. When α is small, we relatively trust the new information more. If $\alpha = 0$, the calibration problem reduces to a simple NLS. There are several approaches to compute the regularization parameter α a posteriori, but Cont and Tankov (2004a, b) propose a use of discrepancy principle by Morozov (1966) which is the oldest, very popular, and a simple posteriori choice rule for α. The first step is to run the unregularized NLS of the equation (15) and obtain the unregularized risk-neutral diffusion parameter estimate $\hat{\sigma}_{\alpha=0}$. The second step is to compute the model intrinsic *a priori* quadratic pricing error $\hat{e}_{\alpha=0}{}^2$ by running the unregularized NLS of the equation (15) with the fixed $\hat{\sigma}_{\alpha=0}$. Let $e^2(\alpha, \theta_\alpha)$ be a posteriori model intrinsic quadratic pricing error for a given regularization parameter $\alpha > 0$ with $\theta_\alpha = (\sigma_\alpha, \lambda_\alpha, \mu_\alpha, \delta_\alpha)$. We expect:

$$e^2(\alpha, \theta_\alpha) > \hat{e}_{\alpha=0}{}^2 ,$$

because of the addition of the relative entropy. Morozov (1966) is willing to trade off the precision of the calibration for the numerical stability and the uniqueness of the calibration solution within the error bounds of the model intrinsic *a priori* quadratic pricing error $\hat{e}_{\alpha=0}{}^2$, thus, a regularization parameter α can be estimated by numerically solving the following equation with gradient-descent algorithms:

$$e^2(\alpha, \theta_\alpha) = c\hat{e}_{\alpha=0}{}^2 , \tag{25}$$

where $c = 1.2$, for example. The final regularized calibration result $\theta^{\mathbb{Q}}$ can be obtained by numerically solving the following optimization problem with $\hat{\alpha}$ using gradient-descent algorithms:

$$\theta^{\mathbb{Q}} = \arg\min_{\theta^{\mathbb{Q}}} \sum_{i=1}^{N} \left| C_i^{\text{model}}(\theta^{\mathbb{Q}}) - C_i^{\text{market}} \right|^2 + \hat{\alpha}\mathcal{E}_T(\mathbb{Q}|\mathbb{P}) . \tag{26}$$

[5] Empirical Example: Obtaining the Prior $\mathbb{P}$

The first and probably the most important step toward the implementation regularized calibration with relative entropy is to obtain the prior probability measure $\mathbb{P}$ because the addition of the relative entropy $\mathcal{E}_T(\mathbb{Q}|\mathbb{P})$ means the introduction of the bias of the risk-neutral martingale measure $\mathbb{Q}$ to the prior measure $\mathbb{P}$. In this section, two very different ways of obtaining the prior $\mathbb{P}$ are implemented using index options.

[5.1] Computation of the Statistical Prior with Time-Series Data

Our data consist of daily closing prices of the futures contract on the S&P 500 index with March 2005 maturity obtained from Chicago Mercantile Exchange (CME) Daily Bulletin for the period March 24, 2004, through March 17, 2005 for the total of 248 trading days. Thus, we have a sample of log return series $(R_{t=1,2,\dots,N})$ of size $N = 247$ where the log return is defined as:

$$R_t = \ln\left(\frac{F_t}{F_{t-1}}\right),$$

where F_t is the futures price.

According to Table 2, the true daily probability distribution of the log return of S&P 500 futures price for this sample shows a slight negative skewness -0.01644 and almost zero excess kurtosis of -0.0003468. True distribution is plotted in Figure 5 using the kernel density estimator with the Gaussian kernel with the bandwidth 0.002:

$$\mathbb{P}_N(R) = \frac{1}{0.002N}\sum_{t=1}^{N}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{R-R_t}{0.002}\right)^2\right\}. \tag{27}$$

Anderson-Darling (AD) test developed by Stephens (1974) is employed to perform the goodness-of-fit test of the following hypotheses:

H_0 : The sample of log return $(R_{t=1,2,\dots,N})$ comes from a population with a normal distribution function .

H_1 : H_0 is not true.

The advantage of AD test over the Kolmogorov-Smirnov test is its sensitivity for the tails of the distribution. AD test statistic is calculated as:

$$A^2 = -N - \frac{1}{N}\left[\sum_{i=1}^{N}(2i-1)\left\{\ln\Phi\left(\frac{R_i-\overline{R}}{s_R}\right)+\ln\left(1-\Phi\left(\frac{R_{N+1-i}-\overline{R}}{s_R}\right)\right)\right\}\right], \tag{28}$$

where Φ is the standard normal cumulative density function, $\bar{R}$ is the sample mean, and s_R is the sample standard deviation. Note that the sample values are rearranged in ascending order:

$$R_1 \leq R_2 \leq \ldots \leq R_N .$$

In our example, the modified AD test statistic is computed as:

$$A^2{*} = A^2 \left(1 + \frac{4}{N} - \frac{25}{N^2} \right) = 1.1412 ,$$

which exceeds the critical value of 0.787 at 5% confidence level. Thus, the normality of the log returns is rejected in this sample.

We use the MLE to estimate the daily probability distribution of the log return under the statistical probability measure $\mathbb{P}$ in the case of the MJD model. MLE is an estimation of the prior model parameter vector $\theta_{\mathbb{P}}$ to maximize the likelihood of observing the particular series of observations. The optimization problem in terms of the log likelihood function is:

$$\max_{\theta_{\mathbb{P}}} l(\theta_{\mathbb{P}}) = \sum_{t=1}^{N} \ln \mathbb{P}(R_t ; \theta_{\mathbb{P}}) ,$$

where $(R_{t=1,2,\ldots,N})$ is the sample log return series and $\mathbb{P}(R_t ; \theta_{\mathbb{P}})$ is the log return probability density given by the equation (12) for the Merton jump diffusion model. Note because these are risk neutral densities, they need to be converted to the statistical density by replacing the instantaneous risk free interest rate r by the instantaneous return on the asset m and the convexity correction $\varpi_{\mathbb{P}}$ is under the statistical probability measure $\mathbb{P}$.

For the estimation of MJD model, small time approximation for the likelihood of increments in the time interval $[t, t+\Delta]$ is employed following Cont and Tankov (2004a):

$$\mathbb{P}(R_\Delta ; \theta) \simeq \lambda \Delta \mathbb{P}(R_\Delta \mid j = 1; \theta) + (1 - \lambda \Delta) \mathbb{P}(R_\Delta \mid j = 0; \theta) , \tag{29}$$

where $\mathbb{P}(R_\Delta \mid j = 1; \theta)$ is the conditional probability density of log return R in a small time Δ given that one jump has occurred and $\mathbb{P}(R_\Delta \mid j = 0; \theta)$ is the conditional probability density of log return R in a small time Δ given that jump did not occur.

The results are reported in Table 1 and 2. Figure 5 gives the estimated log return probability density plot and Figure 6 provides the plot of the log probability density to better illustrate the tail behavior. As expected, the true log return distribution is characterized by the higher peak, the heavier lower tail, and the thinner upper tail than the

BS density. These important features of the true log return probability density are captured better by the MJD model. Similar to Honoré (1998) and Cont and Tankov (2004a), the large estimated value of the intensity parameter in the MJD model $\lambda_{\mathbb{P}} = 60$ (this means that on average the process jumps 60 times per year) casts a doubt to the legitimacy of modeling jumps as a rare event under the statistical measure $\mathbb{P}$. This large estimated value of the intensity parameter in the MJD model suggests that the log return process actually moves by frequent jumps instead of diffusion process and therefore this is an indication to resort to the Lévy models of infinite activity.

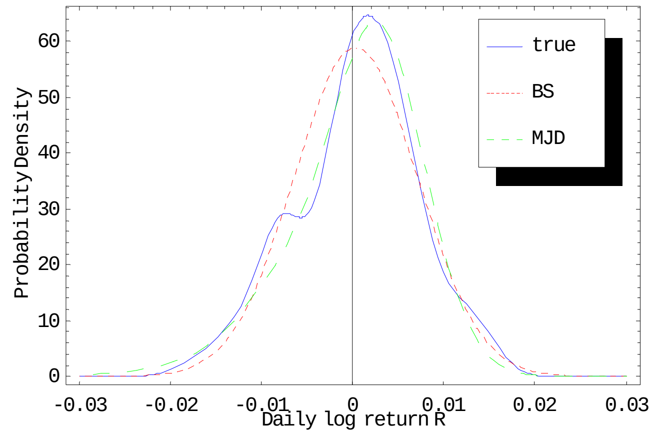


Figure 5 Statistical daily log return probability density $\mathbb{P}(R_t)$

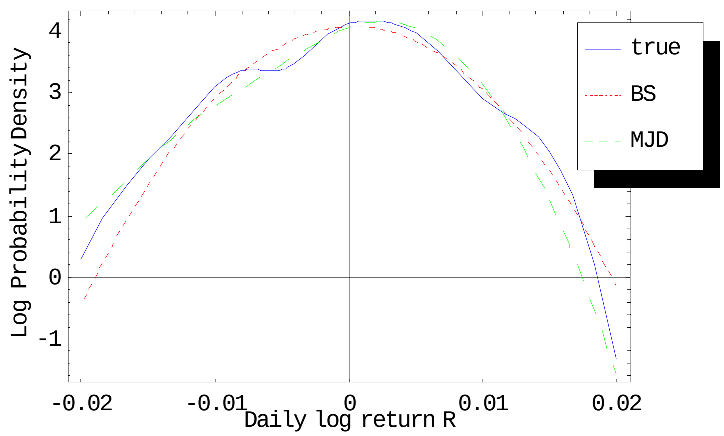


Figure 6 Log of Statistical daily log return probability density $\mathbb{P}(R_t)$

Table 1 Statistically estimated parameters of each model

Model	Log likelihood	Model parameter vector $\theta_{\mathbb{P}}$		
BS	883.466	$\sigma_{\mathbb{P}} = 0.107419$		
MJD	886.568	$\sigma_{\mathbb{P}} = 0.080850$	$\lambda_{\mathbb{P}} = 60.00000$	$\mu_{\mathbb{P}} = -0.010476$
		$\delta_{\mathbb{P}} = 1.600779 \times 10^{-9}$		

Table 2 Annualized standardized moments of statistical log return probability density $\mathbb{P}(R_t)$

Model	Standard Deviation	Skewness	Excess Kurtosis
True	0.107640	-0.016440	-0.000347
BS	0.107419	0	0
MJD	0.114548	-0.045892	0.004197

Next, Figure 7 plots the statistically estimated MJD Lévy measure of the equation (5) using the reported values in Table 1 for the range of $[\mu_{\mathbb{P}} - 10 \times 10^{-8}, \mu_{\mathbb{P}} + 10 \times 10^{-8}]$. The reason that this MJD Lévy measure is a spike at $\mu_{\mathbb{P}}$ is due to the extremely small value of the estimated standard deviation parameter of the log return jump size $\delta_{\mathbb{P}}$.

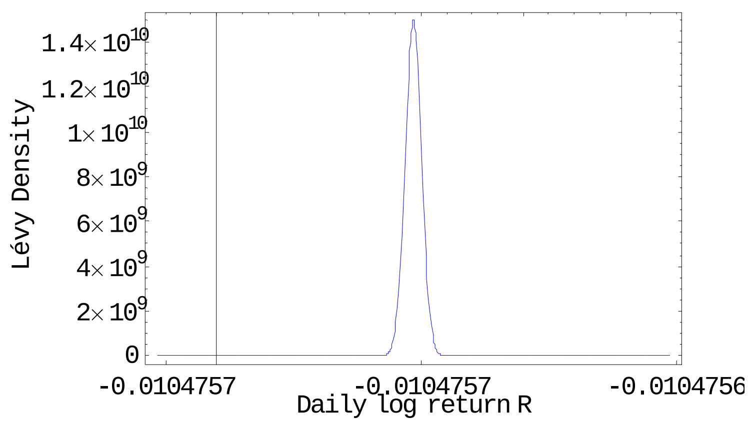


Figure 7 Plot of the Statistically Estimated MJD Lévy Measure $\ell_{MJD,R}(R; \lambda_{\mathbb{P}}, \mu_{\mathbb{P}}, \delta_{\mathbb{P}})$

For a notational clarity, let $\mathbb{P}^S$ be the statistical prior probability measure and $\theta_{\mathbb{P}}^S$ be the statistical prior parameter vector which is:

$$\theta_{\mathbb{P}}^S = (\sigma_{\mathbb{P}}^S = 0.08085, \lambda_{\mathbb{P}}^S = 60, \mu_{\mathbb{P}}^S = -0.010476, \delta_{\mathbb{P}}^S = 1.600779 \times 10^{-9}), \qquad (30)$$

where the superscript 'S' means 'statistical'.

[5.2] Computation of the Risk-Neutral Prior with Option Data

Our data consist of daily settlement call option prices on S&P 500 futures with March 2005 maturity obtained from Chicago Mercantile Exchange (CME) Daily Bulletin for the sample period from March 24, 2004, through March 16, 2005 for the total of 248 trading days.[6] These American style option prices are converted to European style option prices using Barone-Adesi and Whaley (1987) quadratic approximation method to adjust for the early exercise premium. After eliminating call prices less than 0.125 due to reliability issues, the data used consist of a total of 6567 call prices. Daily series of three month Treasury Bill rate are used as appropriate risk-free interest rates.

Using the above data, simply running an unregularized calibration of the equation (15) on each day for the sample period produces the successively updated risk-neutral prior parameter solutions. Let $(\mathbb{P}_t^{RN})$ be the series (i.e. $t = 1, 2, ..., 248$) of successively updated risk-neutral prior probability measures and $(\theta_{\mathbb{P},t}^{RN})$ be the series of successively updated risk-neutral prior parameter vectors which are illustrated in Figure 8.

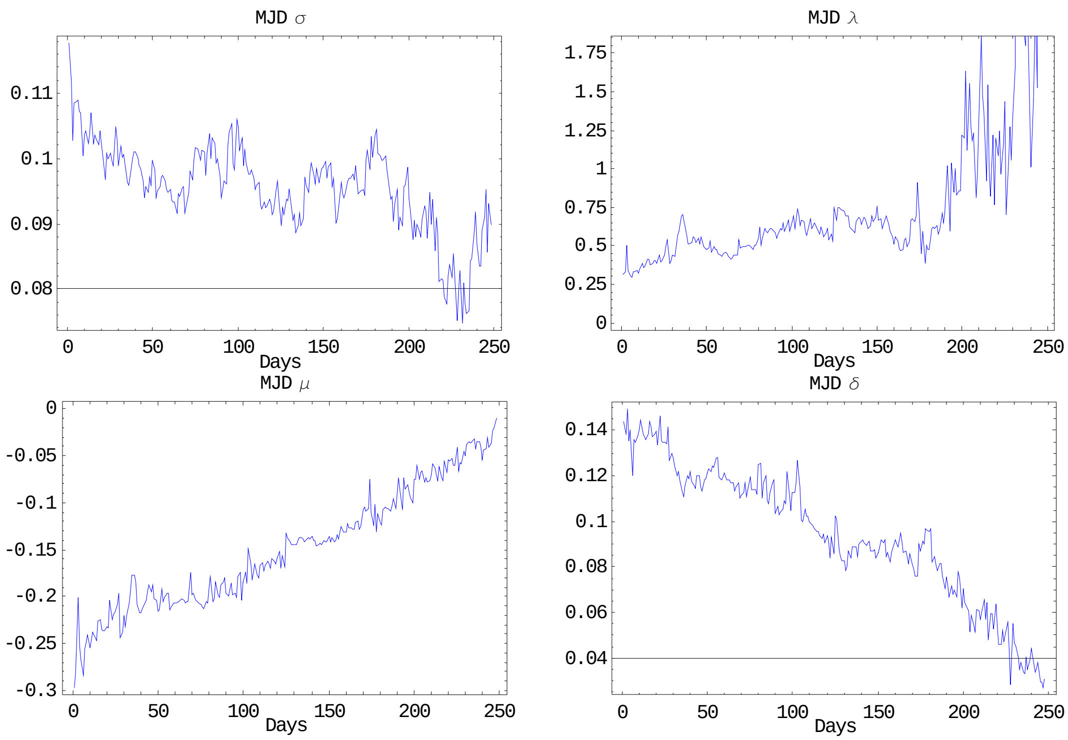


Figure 8 Successively updated risk-neutral prior parameters Time series of calibrated parameters without regularization for the MJD model are plotted from the day 1 which is March 24, 2004, through day 248 which is March 16, 2005.

Taking the average of the series of successively updated risk-neutral prior parameter vectors $(\theta_{\mathbb{P},t}^{RN})$ yields the (fixed) risk-neutral prior parameter vector $\theta_{\mathbb{P}}^{RN}$ and let $\mathbb{P}^{RN}$ be

[6] March 16, 2005 is one day before the last trading day.

the (fixed) risk-neutral prior probability measure. Table 3 reports the MJD risk-neutral prior parameter vector as follows:

$$\theta_{\mathbb{P}}^{RN} = (\sigma_{\mathbb{P}}^{RN} = 0.09544, \lambda_{\mathbb{P}}^{RN} = 0.77742, \mu_{\mathbb{P}}^{RN} = -0.14899, \delta_{\mathbb{P}}^{RN} = 0.09411). \qquad (31)$$

Table 3
Risk-neutral prior parameters

The average, the standard error, the minimum, and the maximum of the successively updated risk-neutral parameters for the sample period are presented.

Parameters		Average	Standard Error	Minimum	Maximum
MJD	σ	0.09544	0.00669	0.07470	0.11761
	λ	0.77742	0.59858	0.29551	5.66366
	μ	-0.14899	0.06361	-0.29654	-0.00985
	δ	0.09411	0.02968	0.02681	0.14922
BS	σ	0.14437	0.02253	0.09714	0.19339

Figure 9 gives the plot of the risk-neutral log return probability density with the maturity $T = 0.25$ years and Figure 10 provides the plot of the log of Figure 9 to better illustrate the tail behavior. As expected, the MJD model captures the negative skewness and the excess kurtosis of the log return probability density under the risk-neutral probability measure.

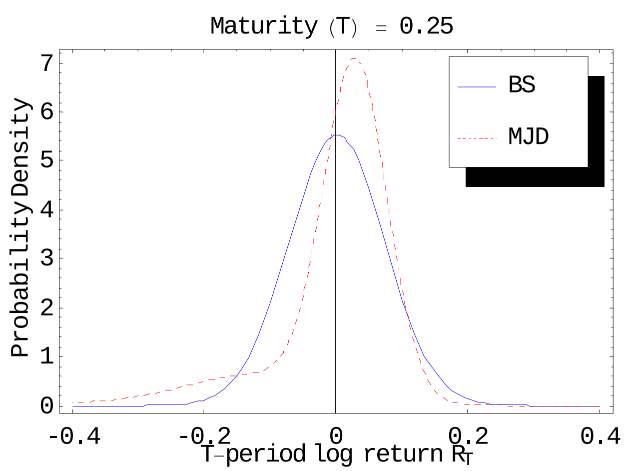


Figure 9 Risk-Neutral log return probability density The maturity is set at $T = 0.25$.

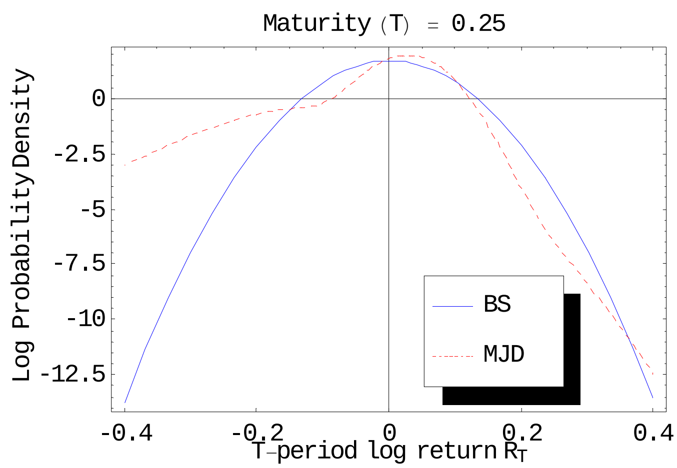


Figure 10 Log of Risk-Neutral daily log return probability density

Table 3 Annualized standardized moments of risk-neutral log return probability density

Model	Standard Deviation	Skewness	Excess Kurtosis
BS	0.14437	0	0
MJD	0.18235	-0.931611	1.34135

Next, Figure 11 plots the risk-neutral MJD Lévy measure of the equation (5) using the reported values in Table 3. It is symmetric around the mean log return jump size $\mu_{\mathbb{P}}^{RN}$ and its total mass is equal to the intensity $\lambda_{\mathbb{P}}^{RN}$. When Figures 7 and 11 are compared, we notice the remarkable difference of the Lévy measure under the statistical probability measure and the risk-neutral probability measure.

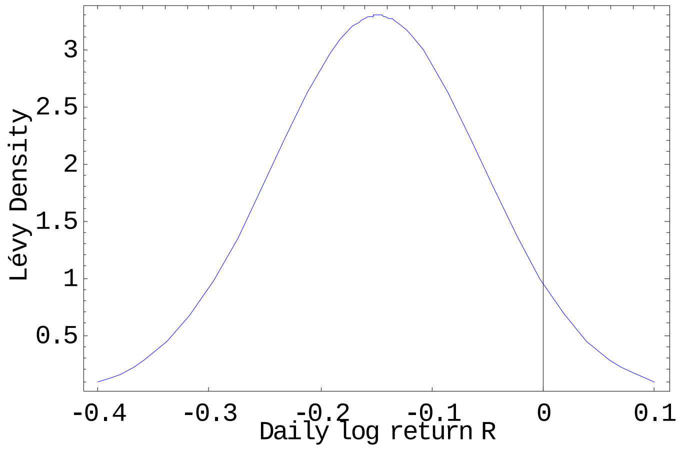
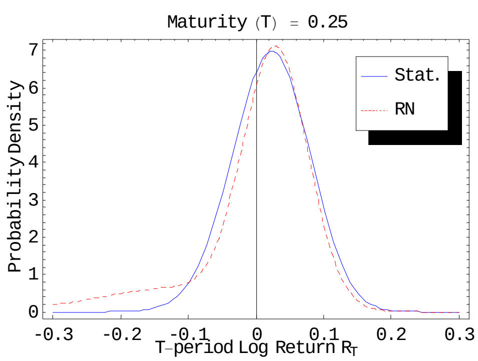


Figure 11 Plot of the risk-neutral MJD Lévy Measure

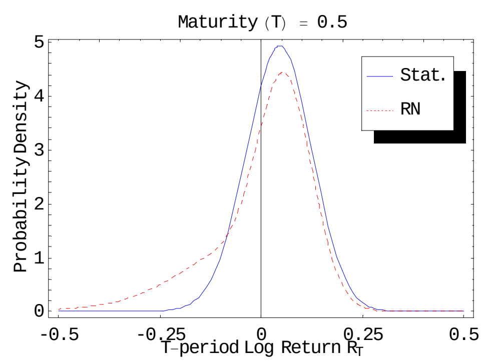
[5.3] Comparison between the Statistical Prior and the Risk-Neutral Prior

Figure 12 illustrates the difference between the statistical prior probability measure $\mathbb{P}^{S}$ and the risk-neutral prior probability measure $\mathbb{P}^{RN}$ for four different maturities $T = 0.25$, 0.5, 0.75, and 1. Table 4 reports the difference in the standardized moments. We observe that the volatility, the negative skewness and the excess kurtosis all become more pronounced under risk-neutral measure $\mathbb{P}^{RN}$ due to the heavier lower tails. This is a well-documented fact which is due to the market participants' fear for the market crush.

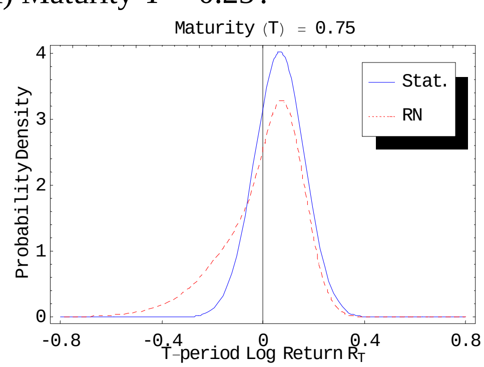
Notice also the remarkable difference between the statistical prior parameter vector $\theta_{\mathbb{P}}^{S}$ of the equation (30) and the risk-neutral prior parameter vector $\theta_{\mathbb{P}}^{RN}$ of the equation (31). Under the risk-neutral probability measure, the asset price jumps only less than once (i.e. 0.77742 times) compared to 60 jumps under the statistical measure. This indicates that modeling jumps as rare events is legitimate under the pricing measure, but not under the statistical measure. The risk-neutral mean log return jump size $\mu_{\mathbb{P}}^{RN}$ is -14.899% which is far larger than the statistical counterpart $\mu_{\mathbb{P}}^{S}$ of -1.0476%. The uncertainty regarding the log return jump size is also much larger under the risk-neutral measure with $\delta_{\mathbb{P}}^{RN} = 0.09411$ compared to the statistical counterpart of $\delta_{\mathbb{P}}^{S} = 1.600779 \times 10^{-9}$.



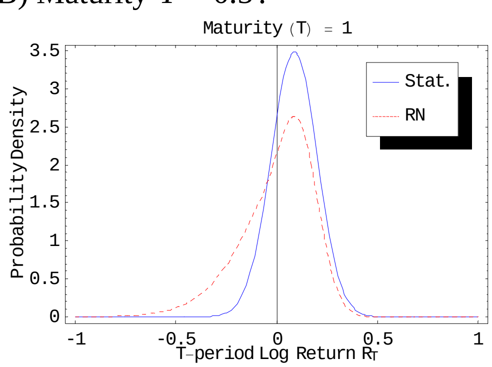
A) Maturity $T = 0.25$.



B) Maturity $T = 0.5$.



C) Maturity $T = 0.75$.



D) Maturity $T = 1$.

Figure 12 Statistical vs. Risk-Neutral prior log return probability densities The statistical prior log return probability density is compared to the risk-neutral prior log return probability density for four different maturities $T = 0.25$, 0.5, 0.75, and 1.

Table 4 Annualized standardized moments of statistical vs. risk-neutral prior log return probability density

Model	Standard Deviation	Skewness	Excess Kurtosis
Statistical	0.114548	-0.045892	0.004197
Risk-neutral	0.18235	-0.931611	1.34135

[6] Empirical Result of Regularized Calibration with Relative Entropy

In this section, we apply Cont and Tankov's (2004 a, b) regularization method with the relative entropy described briefly in the previous section to the index options parametrically with the MJD model.

[6.1] Methods

We compare the result of the calibration under the following three different methods.

Method 1 (M1): Unregularized calibration of the equation (15).
Method 2 (M2): Regularized calibration of the equation (24) with the statistical prior of the equation (30).
Method 3 (M3): Regularized calibration of the equation (24) with the (fixed) risk-neutral prior.

[6.2] Empirical Results

We describe the implemented regularization procedures step by step using the data on March 30, 2004 (245 days to maturity).

Method 1 (M1)

This is the simplest and fastest in computational time and the calibrated risk-neutral parameter vector is shown in Table 5. The sum of the squared pricing errors (the value of the objective function) is 0.132312.

Method 2 (M2)

The first step is to run the unregularized NLS which is the Method 1. We obtain the unregularized risk-neutral diffusion parameter estimate:

$$\hat{\sigma}_{\alpha=0} = 0.108775 .$$

The second step is to compute the model intrinsic *a priori* quadratic pricing error $\hat{e}_{\alpha=0}{}^{2}$ by running the unregularized NLS (i.e. Method 1) with $\hat{\sigma}_{\alpha=0} = 0.108775$. We obtain:

$$\hat{e}_{\alpha=0}{}^{2} = 0.132311 .$$

The third step is to estimate the regularization parameter α by trading off the precision of the calibration for the numerical stability and the uniqueness of the calibration solution within the error bounds of the model intrinsic *a priori* quadratic pricing error $\hat{e}_{\alpha=0}{}^{2}$. In other words, the third step is to numerically solve the following equation with gradient-descent algorithms:

$$e^2(\alpha,\theta_\alpha) = \min_{\theta_\alpha^{\mathbb{Q}}} \sum_{i=1}^{N} \left| C_i^{\text{model}}(\theta_\alpha^{\mathbb{Q}}) - C_i^{\text{market}} \right|^2 + \alpha \mathcal{E}_T(\mathbb{Q}|\mathbb{P}) \approx 1.2 \times 0.132311. \qquad (32)$$

But, the problem is that the equation (32) has no solution on $\mathbb{R}$. The cause is the remarkable difference between the statistical prior parameter vector $\theta_{\mathbb{P}}^S$ of the equation (30) and the unregularized risk-neutral solution. Suppose that the solution of the regularized calibration problem is close to the unregularized risk-neutral solution. Using the statistical prior $\theta_{\mathbb{P}}^S$ of the equation (30) and the result of Method 1 as the solution of the regularized calibration problem, the relative entropy turns out to be too large because the statistical measure and the risk-neutral measure is too different:

$$\mathcal{E}_T(\mathbb{Q}^{M1}|\mathbb{P}^S) = 4.99078 \times 10^{15}.$$

In this case, the equation (32) becomes:

$$\min_{\theta_\alpha^{\mathbb{Q}}} \sum_{i=1}^{N} \left| C_i^{\text{model}}(\theta_\alpha^{\mathbb{Q}}) - C_i^{\text{market}} \right|^2 + \alpha \times 4.99078 \times 10^{15} \approx 1.2 \times 0.132311. \qquad (33)$$

The solution of the regularization parameter α in the equation (33) is approximately zero. Therefore, there is no meaning to regularize the calibration problem. But, this is because of the assumption that the solution of the regularized calibration problem is close to the unregularized risk-neutral solution. Next, consider the opposite case in which the solution of the regularized calibration problem is close to the statistical prior $\theta_{\mathbb{P}}^S$ of the equation (30). Suppose that the solution of the regularized calibration problem is equal to the statistical prior $\theta_{\mathbb{P}}^S$ of the equation (30) for simplicity. The relative entropy is zero by definition:

$$\mathcal{E}_T(\mathbb{P}^S|\mathbb{P}^S) = 0.$$

Thus, the value of the regularization parameter α does not really matter and the sum of the squared pricing errors is calculated as:

$$\sum_{i=1}^{N} \left| C_i^{\text{model}}(\mathbb{P}^S) - C_i^{\text{market}} \right|^2 = 2834.1.$$

Again, this is because of the significant difference between the statistical measure and the pricing measure. This example illustrates the fact that using the statistical prior $\theta_{\mathbb{P}}^S$ of the equation (30), the regularized calibration solution cannot be close to the unregularized risk-neutral solution nor the statistical prior $\theta_{\mathbb{P}}^S$. In addition, there is no regularized calibration solution which is in-between the unregularized risk-neutral solution and the statistical prior $\theta_{\mathbb{P}}^S$ because of the remarkable difference between these two. Numerically speaking, this is equivalent to stating that using the statistical prior $\theta_{\mathbb{P}}^{S}$, the equation (32) has no real-valued solution with constraints:

$$\alpha, \sigma, \lambda, \text{ and } \delta > 0 .$$

Thus, the regularized calibration with the statistical prior is not implementable.

Method 3 (M3)

The first two steps are the same as Method 2. The third step is to numerically solve the equation (32) which is solvable this time because of the proximity between the risk-neutral prior $\mathbb{P}^{RN}$ and the regularized calibration solution (these are both risk-neutral measures). Numerically solving the equation (32) yields the estimate for the regularization parameter as:

$$\hat{\alpha} = 0.043162 .$$

Finally, regularized calibration result $\theta^{\mathbb{Q}}$ shown in Table 5 can be obtained by numerically solving the following optimization problem with $\hat{\alpha}$ using gradient-descent algorithms:

$$\theta^{\mathbb{Q}} = \arg\min_{\theta^{\mathbb{Q}}} \sum_{i=1}^{N} \left| C_i^{\text{model}}(\theta^{\mathbb{Q}}) - C_i^{\text{market}} \right|^2 + 0.043162 \times \mathcal{E}_T(\mathbb{Q}|\mathbb{P}) .$$

The sum of the squared pricing errors (the value of the objective function) is 0.156149:

$$\sum_{i=1}^{N} \left| C_i^{\text{model}}(\theta^{\mathbb{Q}}) - C_i^{\text{market}} \right|^2 + 0.043162 \times \mathcal{E}_T(\mathbb{Q}|\mathbb{P}) = 0.156149 ,$$

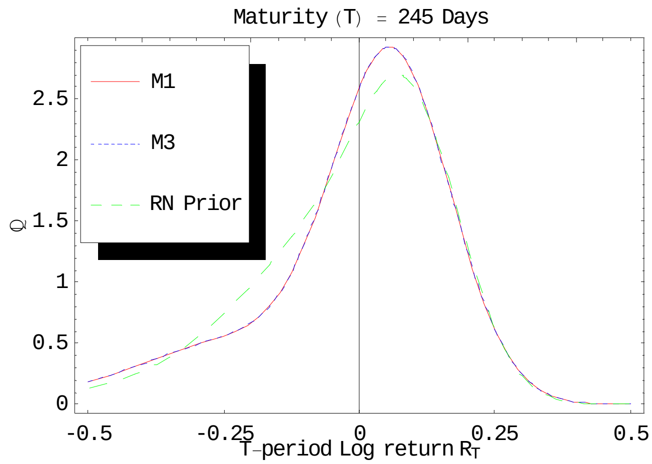
where $\sum_{i=1}^{N} \left| C_i^{\text{model}}(\theta^{\mathbb{Q}}) - C_i^{\text{market}} \right|^2 = 0.132777$ and $\mathcal{E}_T(\mathbb{Q}|\mathbb{P}) = 0.541481$.

Table 5 reports the results of the calibration, and Figure 13 through 18 compare the calibrated log return probability density and the calibrated Lévy measure for each method on six different maturity dates. We observe that the calibrated risk-neutral parameters are quite similar with or without regularization. This can be confirmed by observing the almost same calibrated log return probability densities and the Lévy measures between Method 1 and Method 3. It seems that Lévy measures are more sensitive to the regularization (i.e. the small difference in the calibrated parameters) than log return probability measures which is shown in Panel C of Figure 18. We also find that the difference in calibrated parameters between the regularized and the unregularized become more pronounced especially for near maturity options shown in Figure 18.

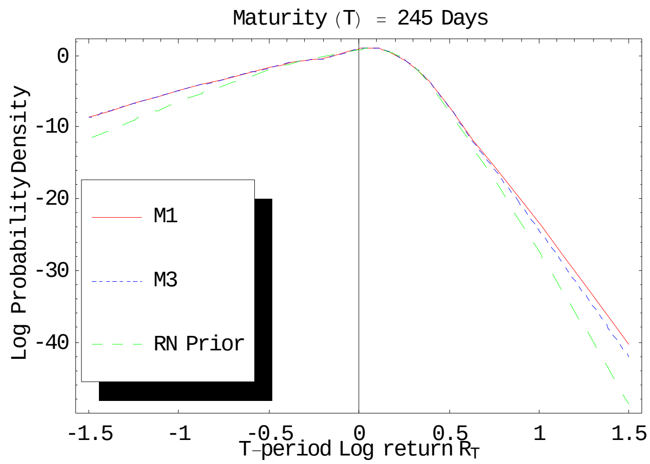
We believe that the reason of this no significant difference between the unregularized calibration and the regularized calibration is the use of $c = 1.2$ in the equation (25) and (32). This constant c controls the degree of trading off the precision for the numerical stability and uniqueness of the calibration. The larger c indicates more willingness to sacrifice the precision and $c = 1$ means that the calibration is unregularized (i.e. $\alpha = 0$). It is our opinion that the use of $c = 1.2$ does not sacrifice the precision too much for the numerical stability and uniqueness which means that the regularized calibration result will by design be close to the unregularized calibration result. But, we are not suggesting the use of larger c (for example 2) because it can be a too much of sacrifice of the precision. So, what is the optimal value for c? Cont and Tankov (2004a, b) recommend the use of $1.1 < c < 1.5$. But, in the end, the choice of c is entirely up to the user's discretion. One thing to remind is that the larger c means the introduction of larger bias toward the prior and the regularized calibration result will show more significant difference from the result of the unregularized calibration. But, the use of larger c means the user's stronger belief in the prior information than the today's information. Such situation is very difficult to imagine.

Table 5
Calibration Results

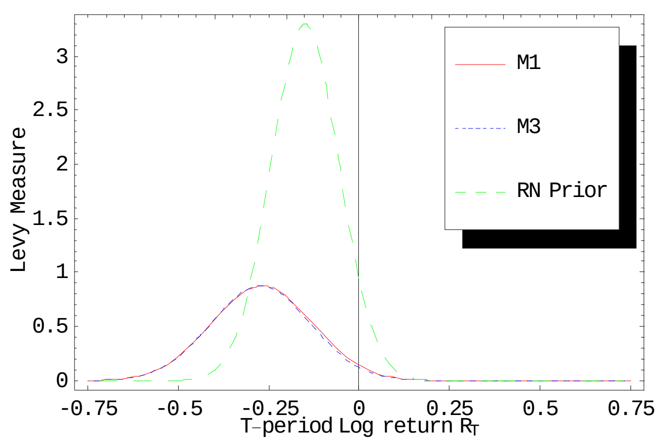
Methods	$\sigma^{\mathbb{Q}}$	$\lambda^{\mathbb{Q}}$	$\mu^{\mathbb{Q}}$	$\delta^{\mathbb{Q}}$
March 30, 2004 (245 days to maturity)				
M1	0.108775	0.307107	-0.267493	0.140028
M2				
M3	0.108995	0.300333	-0.272398	0.136964
July 1, 2004 (180 days to maturity)				
M1	0.093676	0.485600	-0.197655	0.111576
M2				
M3	0.092968	0.509876	-0.190205	0.114493
September 27, 2004 (120 days to maturity)				
M1	0.095187	0.732404	-0.144411	0.082790
M2				
M3	0.095463	0.714316	-0.147262	0.081582
November 22, 2004 (80 days to maturity)				
M1	0.094648	0.662536	-0.107062	0.080695
M2				
M3	0.095101	0.630734	-0.111636	0.079383
January 20, 2005 (40 days to maturity)				
M1	0.091795	1.296250	-0.070378	0.056700
M2				
M3	0.091314	1.259500	-0.075242	0.052736
March 4, 2005 (10 days to maturity)				
M1	0.086929	1.009830	-0.054321	0.043944
M2				
M3	0.088270	0.736562	-0.067053	0.046267



A) Calibrated log return probability density.

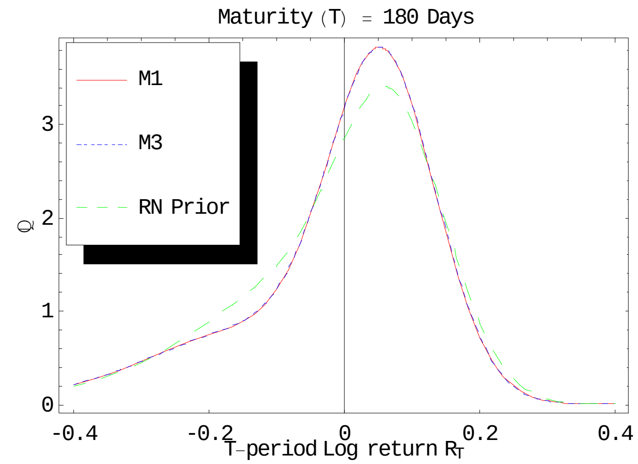


B) Log of Panel A).

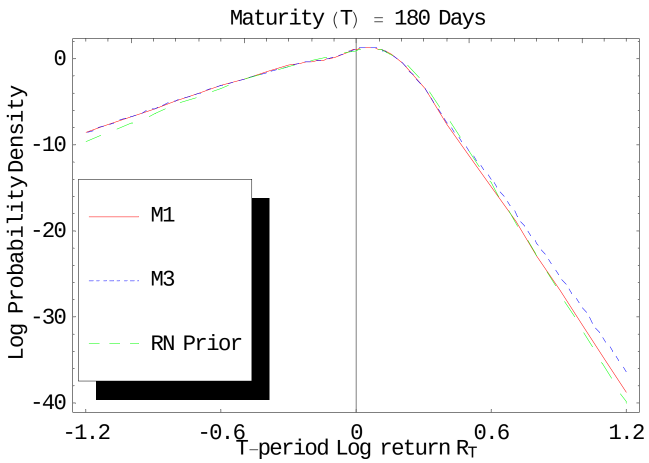


C) Calibrated Lévy measure.

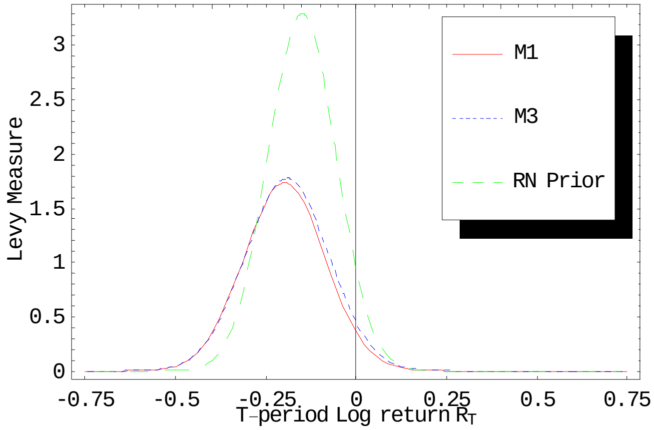
Figure 13 Plot of calibrated log return probability density and Lévy measure on March 30, 2004 (245 days to maturity)



A) Calibrated log return probability density.

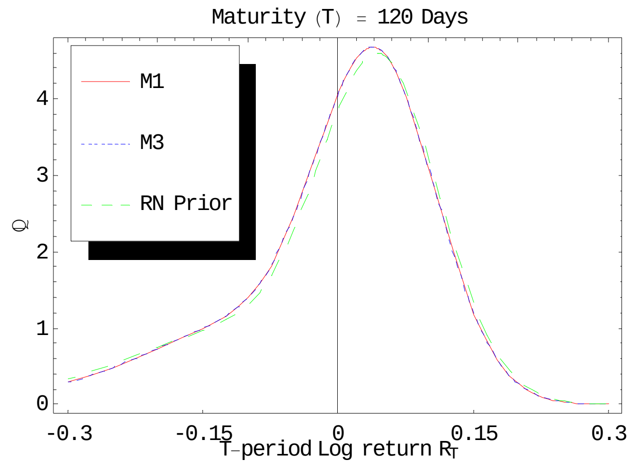


B) Log of Panel A).

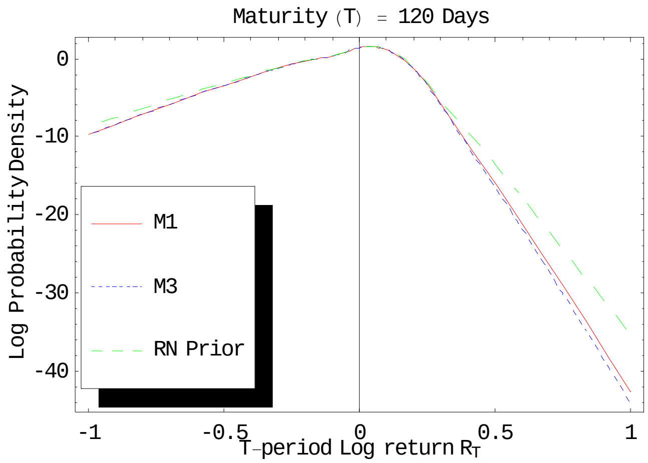


C) Calibrated Lévy measure.

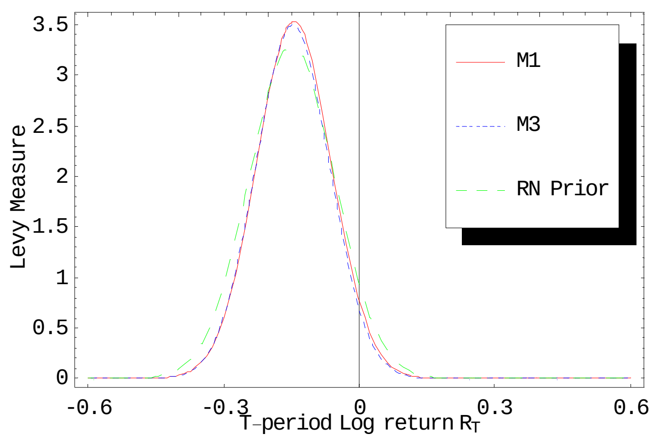
Figure 14 Plot of calibrated log return probability density and Lévy measure on July 1, 2004 (180 days to maturity)



A) Calibrated log return probability density.

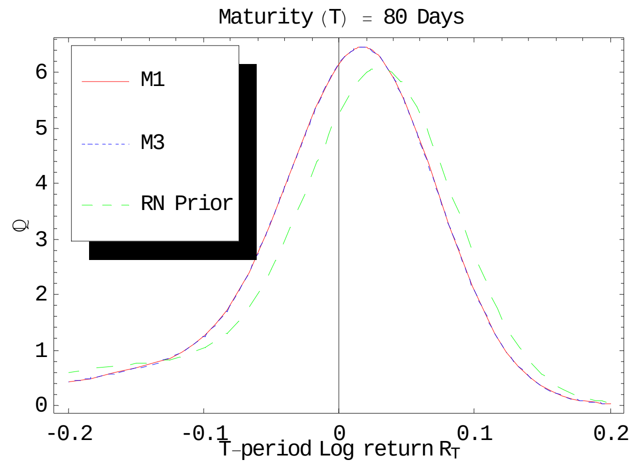


B) Log of Panel A).

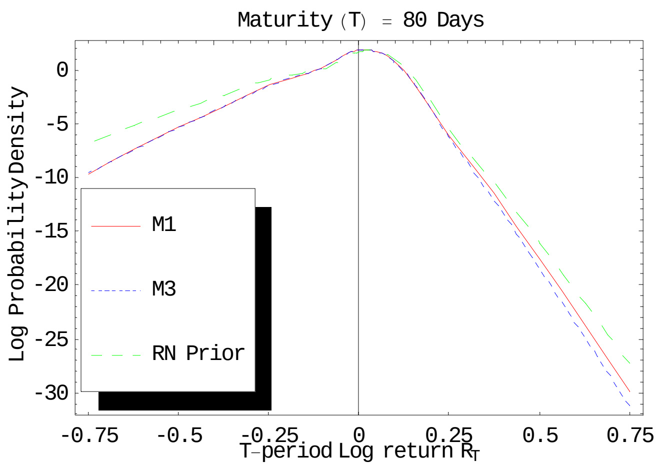


C) Calibrated Lévy measure.

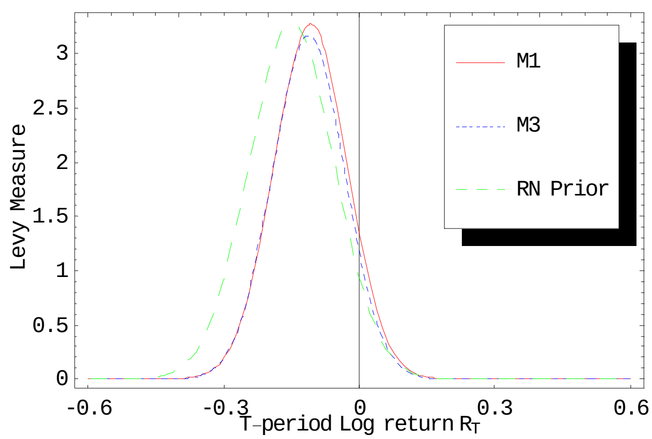
Figure 15 Plot of calibrated log return probability density and Lévy measure on September 27, 2004 (120 days to maturity)



A) Calibrated log return probability density.

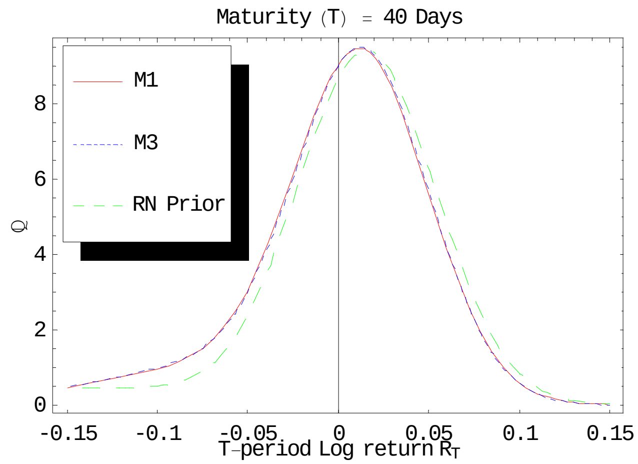


B) Log of Panel A).

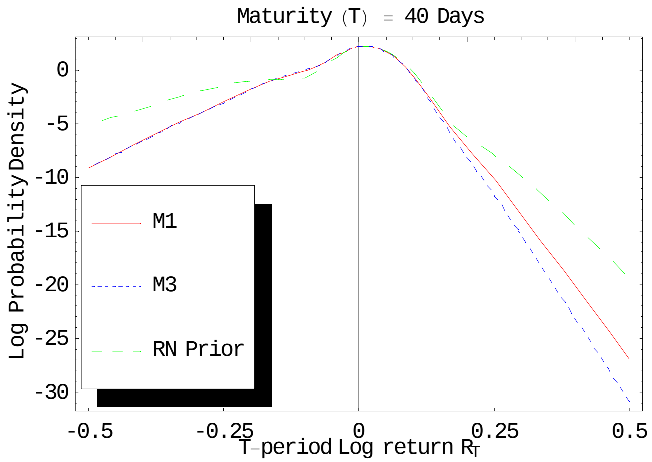


C) Calibrated Lévy measure.

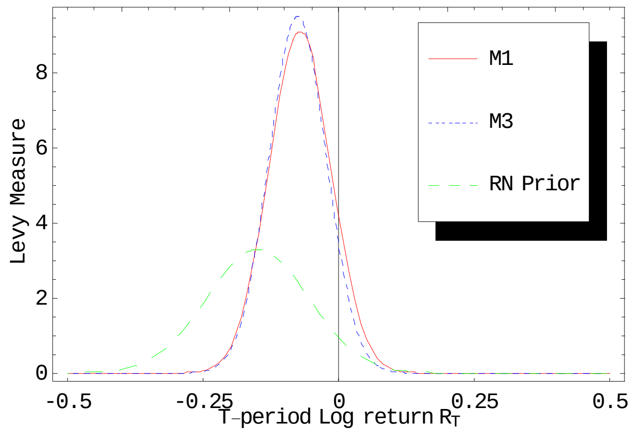
Figure 16 Plot of calibrated log return probability density and Lévy measure on November 22, 2004 (80 days to maturity)



A) Calibrated log return probability density.

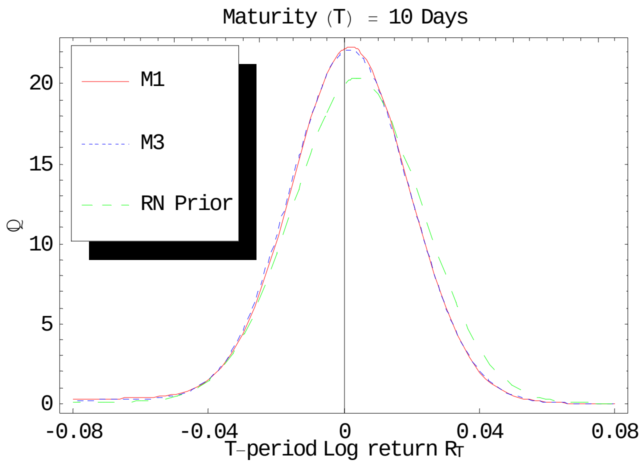


B) Log of Panel A).

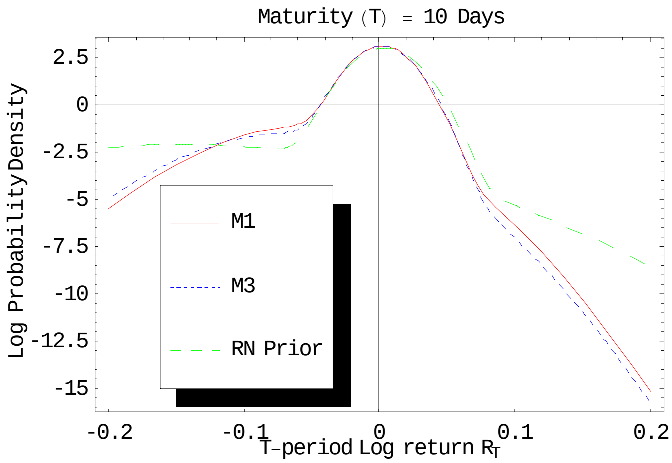


C) Calibrated Lévy measure.

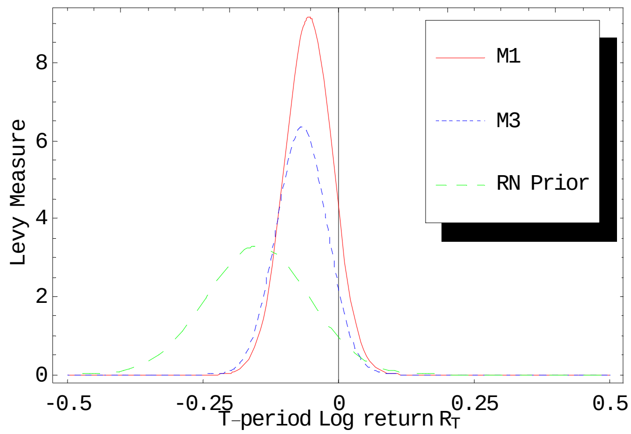
Figure 17 Plot of calibrated log return probability density and Lévy measure on January 20, 2005 (40 days to maturity)



A) Calibrated log return probability density.



B) Log of Panel A).



C) Calibrated Lévy measure.

Figure 18 Plot of calibrated log return probability density and Lévy measure on March 4, 2005 (10 days to maturity)

[7] Conclusion

In this paper, we investigate the effect of the regularization with the relative entropy in the framework of the parametric calibration of the Merton jump-diffusion model to the index options. It is important to realize that using the relative entropy and using the prior mean the introduction of the bias of the calibration solution toward the prior to gain the numerical stability and the uniqueness by making the objective function more convex. This means that the user has some belief in the use of the prior (i.e. otherwise, why do you bother?). Another important point is that the regularization is not the method to locate the global solution, it is the method to enhance the uniqueness and the stability of the calibration solution by sacrificing its precision. In terms of the choice of the prior, the only implementable prior is the risk-neutral prior. The regularized calibration with the statistical prior is not implementable because the statistical prior is a statistical measure and it is too much away from the risk-neutral (i.e. pricing) measure.

The result shows that with or without regularization, calibrated risk-neutral parameters, calibrated log return probability densities, and calibrated Lévy measures are not significantly different. It seems that Lévy measures are more sensitive to the regularization than log return probability measures. Notice also that the difference in calibrated parameters between the regularized and the unregularized become more pronounced especially for near maturity options.

From our empirical result that the regularization with the relative entropy using the risk-neutral prior does not make any significant difference in the solution and considering the extra computation time necessary for the regularization procedures, we prefer the calibration without any regularization which uses only today's information and yields the unbiased solution although the unregularized calibration solution may not be stable and unique.

References

Avellaneda, M., Friedman, C., Holmes, R., and D. Samperi. “Calibrating Volatility Surfaces via Relative-Entropy Minimization.” *Applied Mathematical Finance*, 4 (1997), 37-64.

Bakshi, G., Cao, C., and Z. Chen. “Empirical Performance of Alternative Option Pricing Models.” *Journal of Finance*, 52 (1997), 2003-49.

Bates, D. “Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutschemark Options.” *Review of Financial Studies*, 9 (1996), 69-108.

Black, F., and M. Scholes. “The Pricing of Options and Corporate Liabilities.” *Journal of Political Economy*, 81 (May/June 1973), 637-59.

Carr, P., and D. B. Madan. “Option Valuation Using the Fast Fourier Transform.” *Journal of Computational Finance*, 2 (1998), 61-73.

Carr, P., Chang, E. C., and D. B. Madan. “The Variance Gamma Process and Option Pricing.” *European Finance Review*, 2 (1998), 79-105.

Carr, P., Geman, H., Madan, D. B., and M. Yor. “The Fine Structure of Asset Returns: An Empirical Investigation.” *Journal of Business*, 75 (2002), 305-32.

Cont, R., and P. Tankov. *Financial Modelling with Jump Processes*. Chapman & Hall/CRC Financial Mathematics Series, 2004a.

Cont, R., and P. Tankov. “Non-Parametric Calibration of Jump-Diffusion Option Pricing Models.” *Journal of Computational Finance*, 7 (2004b), 1-49.

Cover, T. M., and J.A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1998.

Dumas, B., Fleming, J., and R.E. Whaley. “Implied Volatility Functions: Empirical Tests.” Journal of Finance, 53 (1998), 2059-2106.

Engl, H. W., Hanke, M., and A. Neubauer. *Regularization of Inverse Problems.* Kluwer Academic Publishers Group: Dordrecht, 1996.

Honoré, P. “Pitfalls in Estimating Jump-Diffusion Models.” Working Paper Series, CAF Aarhus School of Business, University of Aarhus, No. 18 (1998).

Matsuda, K. “Introduction to Option Pricing with Fourier Transform: Option Pricing with Exponential Lévy Models.” A part of Ph.D. thesis filed on May 2006, Graduate School and University Center of the City University of New York, 2004.

Matsuda, K. “Introduction to the Mathematics of Lévy Processes.” A part of Ph.D. thesis filed on May 2006, Graduate School and University Center of the City University of New York, 2005a.

Matsuda, K. “Calibration of Lévy Option Pricing Models: Application to S&P 500 Futures Option.” A part of Ph.D. thesis filed on May 2006, Graduate School and University Center of the City University of New York, 2005b.

Merton, R. C. “Option Pricing When Underlying Stock Returns Are Discontinuous.” *Journal of Financial Economics*, 3 (1976), 125-144.

Morozov, V. “On the Solution of Functional Equations by the Method of Regularization.” Soviet Mathematical Doklady, 7 (1966), 414-417.

Sato, K. *Lévy process and Infinitely Divisible Distributions*. Cambridge University Press, 1999.

Stephens, M. A. “EDF Statistics for Goodness of Fit and Some Comparisons.” *Journal of the American Statistical Association*, Vol. 69 (1974), 730-737.