ECOLE POLYTECHNIQUE

CENTRE DE MATHÉMATIQUES APPLIQUÉES
UMR CNRS 7641

91128 PALAISEAU CEDEX (FRANCE). Tél: 01 69 33 41 50. Fax: 01 69 33 30 11

http://www.cmap.polytechnique.fr/

Calibration of jump-diffusion opion pricing models: a robust non-parametric approach.

Rama CONT & Peter TANKOV

R.I. N^0 490 *Oct 2002*

Calibration of jump-diffusion option-pricing models: a robust non-parametric approach.*

Rama Cont & Peter Tankov
Centre de Mathématiques Appliquees
CNRS - Ecole Polytechnique F 91128 Palaiseau, France.
Email: Rama.Cont@polytechnique.org Peter.Tankov@polytechnique.org

October 2002.

Keywords: Lévy process, jump-diffusion models, option pricing, model calibration, non-parametric methods, inverse problems, relative entropy, regularization.

Abstract

We present a non-parametric method for calibrating jump-diffusion models to a finite set of observed option prices. We show that the usual formulations of the inverse problem via nonlinear least squares are ill-posed and propose a regularization method based on relative entropy: we reformulate our calibration problem into a problem of finding a risk neutral jump-diffusion model that reproduces the observed option prices and has the smallest possible relative entropy with respect to a chosen prior model. Our approach allows to conciliate the idea of calibration by relative entropy minimization with the notion of risk neutral valuation in a continuous time model. We discuss the numerical implementation of our method using a gradient based optimization algorithm and show via simulation tests on various examples that the entropy penalty resolves the numerical instability of the calibration problem. Finally, we apply our method to data sets of index options and discuss the empirical results obtained.

*Preliminary versions of this work were presented at the Bernoulli Society International Statistical Symposium (Taipei 2002), Maphysto (Aarhus & Copenhagen), University of Freiburg, University of Warwick and INRIA. We thank Marco Avellaneda, Frédéric Bonnans, Stephane Crépey and Dilip Madan for helpful remarks.

Contents

1 Introduction

The insufficiency of diffusion models to explain certain empirical properties of asset returns and option prices has led to the development, in option pricing theory, of a variety of jump-diffusion models based on Lévy processes[4, 14, 13, 12, 23, 25, 24, 27, 31]. A widely studied class is that of exponential Lévy processes in which the price of the underlying asset is written as $S_t = \exp(rt + X_t)$ where r is the discount rate and X is a Lévy process defined by its characteristic triplet (b, σ, ν) (see section 2.1). While the main concern in the literature has been to obtain efficient analytical and numerical procedures for computing prices of various options, a preliminary step in using the model is to obtain model parameters – here the characteristic triplet of the Lévy process – from market data by calibrating the model to market prices of (liquid) call options. This amounts to solving the following *inverse problem*:

Calibration Problem 1. *Given prices of call options $C_t^*(T_i, K_i), i \in I$, find a Lévy triplet (b, σ, ν) such that the discounted asset price $S_t \exp(-rt)$ is a martingale and the observed option prices are given by their discounted risk neutral expectations:*

$$\forall i \in I, C_t^*(T_i, K_i) = e^{-r(T-t)} E^{(b,\sigma,\nu)}[(S(T_i) - K_i)^+ | S_t = S]. \tag{1}$$

Note that, in order to price exotic options, we need to retrieve the risk neutral *process* and not only its conditional densities (also called the state price densities) as in [1]. Problem (1) is equivalent to a moment problem for the Lévy process X, which is typically an ill posed problem: there may be either no solution at all or an infinite number of solutions. Even in the case where we use an additional criterion to choose one solution from many, the dependence on input prices may be discontinuous, which results in numerical instability of calibration algorithm.

In order to circumvent these difficulties, we propose a *regularization* method based on relative entropy minimization. Our method is based on the idea that, unlike the diffusion setting where different volatility structures lead to singular (non equivalent) measures (and therefore infinite relative entropy), two Lévy processes with different Lévy measures can define equivalent measures. It turns out that the relative entropy of exponential Lévy models is a simple functional of their Lévy measures which can be used as a regularization criterion for solving the inverse problem (1) in stable way. Our approach leads to a nonparametric method for calibrating jump-diffusion models to option prices, extending similar methods previously developed for diffusion models [29].

The paper is structured as follows. Section 2 defines the model set-up and recalls some useful properties of Lévy processes and relative entropy. Section 3 proposes a well-posed formulation of the calibration problem as that of finding a jump-diffusion model that reproduces observed option prices and has the smallest possible relative entropy with respect to some carefully chosen prior measure. Section 4 discusses the numerical implementation of the calibration method, the main ingredient of which is an explicit representation for the gradient of the criterion being minimized (section 4.4).

To assess the performance of our method we first perform numerical experiments on simulated data : calibration is performed on a set of option prices generated from a given exp-Lévy model. Results are presented in section 5: while the non-linear least squares algorithm does not converge in a stable way our algorithm allows to retrieve the Lévy measure while avoiding high sensitivity to the prior. The precision of recovery is especially good for medium and large sized jumps but small jumps are hard to distinguish from a continuous diffusion.

Section 6 presents empirical results obtained by applying our calibration method to a data set of DAX index options. Our tests reveal a density of jumps with strong negative skewness. While a small value of jump intensity seems sufficient to calibrate the observed implied volatility patterns, the shape of the density of jump sizes evolves across maturities, indicating the need for departure from time homogeneity.

2 Model setup

We consider here the class of *exponential Lévy* models where the risk neutral dynamics of the underlying asset is given by $S_t = \exp(rt + X_t)$ where X_t is a (time-homogeneous) jump-diffusion process, also called a *Lévy* process.

2.1 Lévy processes: definitions

A Lévy process is defined as a stochastic process X_t with stationary independent increments which is continuous in probability. Without loss of generality we assume that $X_0 = 0$. The characteristic function of X_t has the following form, called the Lévy-Khinchin representation [30]:

$$E[e^{izX_t}] = \exp\{t(-\frac{1}{2}az^2 + i\gamma_0 z + \int_{-\infty}^{\infty}(e^{izx} - 1 - izx1_{|x|\leq 1})\nu(x)dx)\} \quad (2)$$

where $a > 0$ and γ_0 are real constants and ν is a positive measure verifying

$$\nu(\{0\}) = 0 \qquad \int_{-1}^{+1} x^2\nu(dx) < \infty \qquad \int_{|x|>1} \nu(dx) < \infty \quad (3)$$

We will denote the set of such measures by $\mathcal{L}(\mathbb{R})$. Any Lévy process X can be decomposed into a Brownian motion with drift, a compensated jump process J^1_t with jumps sizes less than or equal to 1 and a jump process J^2 with jumps sizes > 1 [30]:

$$X_t = a\, W_t + \gamma_0 t + J^1_t + J^2_t \quad (4)$$

J^2 (resp. J^1) can then be considered as a superposition of independent Poisson (resp. compensated Poisson) processes with various jump sizes x, $\nu(dx)$ being the intensity (probability per unit time) of jumps of size x. If the measure $\nu(dx)$ admits a density with respect to the Lebesgue measure, we will call it the Lévy density of X and denote its value by $\nu(x)$.

The sample paths of a Lévy process are discontinuous; one may always choose a version of the process such that all sample paths are right continuous with left limits (*càdlàg*). $(X_t, t \in [0,T])$ therefore defines a probability measure of the space of càdlàg functions on $[0,T]$. One can therefore choose Ω to be this space, $\mathcal{F}_t$ to be the corresponding σ-field generated by the paths between 0 and t completed by null sets and $\mathcal{F} = \mathcal{F}_T$.

In general ν is not a probability measure: $\int \nu(dx)$ need not even be finite. In the case where $\lambda = \int \nu(dx) < +\infty$, the Lévy process is said to be of finite activity and the measure ν can then be normalized to define a *probability measure* μ on $\mathbb{R} - \{0\}$ which can be interpreted as the distribution of jump sizes:

$$\mu(dx) \quad = \quad \frac{\nu(dx)}{\lambda} \tag{5}$$

In this case X is called a *compound Poisson* process and λ which is the average number of jumps per unit time, is called the intensity of jumps. In this case the truncation of small jumps is not needed and the Lévy-Khinchin representation reduces to:

$$E[e^{izX_t}] = \exp\{t(-\frac{1}{2}az^2 + i\gamma z + \int_{-\infty}^{\infty}(e^{izx} - 1)\nu(x)dx)\} \tag{6}$$

For further details on Lévy processes see [9, 20, 30].

2.2 Exponential Lévy models

Let $(S_t)_{t\in[0,T^*]}$ be the price of a financial asset modeled as a stochastic process on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{Q})$. Under the hypothesis of absence of arbitrage there exists a measure equivalent to $\mathbb{Q}$ under which $(e^{-rt}S_t)$ is a martingale. We will assume therefore without loss of generality that $\mathbb{Q}$ is already one such martingale measure.

We call *exponential Lévy* model, a model where the dynamics of S_t under $\mathbb{Q}$ is represented as the exponential of a Lévy process:

$$S_t = e^{rt+X_t} \tag{7}$$

Here X_t is a Lévy process with characteristic triplet (σ,γ,ν) and the interest rate r is included for ease of notation. Since the discounted price process $e^{-rt}S_t = e^{X_t}$ is a martingale, this gives a constraint on the triplet (σ,γ,ν):

$$\phi(1) = 0 \iff \gamma = \gamma(\sigma,\nu) = -\frac{\sigma^2}{2} - \int (e^y - 1 - y1_{|y|\leq 1})\nu(dy) \tag{8}$$

We will assume this relation holds in the sequel.

Different exponential Lévy models proposed in the financial modeling literature simply correspond to different parametrizations of the Lévy measure:

• Compound Poisson models: $X_t = \sum_{i=1}^{N^\lambda(t)} Y_i$, $Y_i \sim \nu_0$ IID)
Merton model [27]: $\mu = N(0,\sigma^2)$
Poisson jumps: $\nu = \sum_{k=1}^{n} p_k \delta_{y_k}$.
Double exponential [23] : $\nu(x) = [1_{x>0} p\alpha_1 e^{-\alpha_2 x} + (1-p)\alpha_2 e^{-\alpha_2 x} 1_{x<0}]$
• Variance Gamma [24] $\nu(x) = A|x|^{-1}\exp(-\eta_\pm |x|)$
• Tempered stable[1] processes [22, 12]: $\nu(x) = A_\pm |x|^{-(1+\alpha)}\exp(-\eta_\pm |x|)$
• Normal inverse gaussian process [6]
• Hyperbolic and generalized hyperbolic processes [14, 13]
• Meixner process [31]: $\nu(x) = \frac{Ae^{-ax}}{sinh(x)}$

The price of an option is computed as a discounted conditional expectation of its terminal payoff under the risk-neutral probability Q. By stationarity and independence of increments of X_t, the value of a call option can be expressed as:

$$\begin{aligned} C(t,S;T=t+\tau,K) &= e^{-r\tau}E[(S_T-K)^+|S_t=S] && (9)\\ e^{-r\tau}E[(Se^{r\tau+X_\tau}-K)^+] &= Ke^{-r\tau}E(e^{x+X_\tau}-1)^+ && (10)\end{aligned}$$

Defining the *log forward moneyness* variable

$$x = \ln(S/K) + r\tau \tag{11}$$

one can express the option price via $u(\tau,x) = e^{r\tau}C(t,S;T=t+\tau,K)/K$ which then takes a simpler form:

$$u(\tau,x) = E[(e^{x+X_\tau}-1)^+] = \int \rho(t,dy)(e^{x+y}-1)^+ \tag{12}$$

The pattern of call option prices thus only depends on the current level of underlying and the Lévy triplet $(\sigma,\nu,\gamma(\sigma,\nu))$.

2.3 Equivalence of measures for Lévy processes

One of the interesting properties of models with discontinuous sample paths is that the class of martingale measures equivalent to a given one is quite large. This remains true even of one restricts the price process to remain of exponential-Lévy type under the risk neutral measure. The following result, stated without proof, gives a description of the set of Lévy processes equivalent to a given one. Similar results may be found in [20].

Proposition 1 (Sato [30], Thm 33.1 & 33.2). *Let (X_t,P) and (X_t,P') be two Lévy processes with characteristic triplets are (a,γ,ν) and (a',γ',ν' defined by their corresponding probability measures on the space of càdlàg trajectories. Then $P|_{\mathcal{F}_t}$ and $P'|_{\mathcal{F}_t}$ are mutually absolutely continuous for all t if and only if the three following conditions are satisfied:*

[1]Also called "truncated Lévy flights" in the physics literature.

1. $a = a'$

2. The Lévy measures are mutually absolutely continuous with

$$\int_{-\infty}^{\infty} (e^{\phi(x)/2} - 1)^2 \nu(dx) < \infty \tag{13}$$

where $\phi(x)$ *is defined by* $e^{\phi(x)} = \frac{d\nu'}{d\nu}$

3. If $a = 0$ *then we must in addition have* $\gamma' - \gamma = \int_{-1}^{1} x(\nu' - \nu)(dx)$

The Radon-Nikodym derivative is given by

$$\frac{dP'|_{\mathcal{F}_t}}{dP|_{\mathcal{F}_t}} = e^{U_t} \tag{14}$$

where U_t *is a Lévy process with characteristic triplet*

$$a_U = a\eta^2 \tag{15}$$

$$\nu_U = \nu\phi^{-1} \tag{16}$$

$$\gamma_U = -\frac{1}{2}a\eta^2 - \int_{-\infty}^{\infty} (e^y - 1 - y1_{|y|\leq 1})(\nu\phi^{-1})(dy) \tag{17}$$

and η *is chosen so that*

$$\gamma' - \gamma - \int_{-1}^{1} x(\nu' - \nu)(dx) = a\eta$$

With this choice of drift we have $E^P[e^{U_t}] = 1$

The above result shows an interesting feature of models with jumps compared to diffusion models: we have considerable freedom in changing the Lévy measure, and therefore the option prices, while retaining the equivalence of measures.

Example: tempered stable processes The tempered stable process (also called "truncated" stable processes), introduced by Koponen [22], has a Lévy measure of the following form:

$$\nu(x) = \frac{e^{-\beta^+ x}}{x^{1+\alpha^+}} 1_{x\geq 0} + \frac{e^{-\beta^-|x|}}{|x|^{1+\alpha^-}} 1_{x<0} \tag{18}$$

with $\beta^+ > 0$, $\beta^- > 0$, $0 < \alpha^+ < 2$ and $0 < \alpha^- < 2$. Two tempered stable processes are mutually absolutely continuous if and only if their coefficients α^+ and α^-, which describe the behavior of the Lévy measure near zero, coincide. In fact, the condition (13) for, say, the Lévy measure on the positive half-axis is:

$$\int_0^{\infty} \left(\frac{e^{-\frac{1}{2}(\beta_2^+ - \beta_1^+)x}}{x^{\frac{\alpha_2^+ - \alpha_1^+}{2}}} - 1 \right)^2 \frac{e^{-\beta_1^+ x}}{x^{1+\alpha_1^+}} dx$$

When $\alpha_2^+ < \alpha_1^+$ the integrand is equivalent to $\frac{1}{x^{1+\alpha_1^+}}$ near zero and, hence, is not integrable; the case $\alpha_2^+ > \alpha_1^+$ is symmetric. However, when $\alpha_2^+ = \alpha_1^+$, the integrand is equivalent to $\frac{1}{x^{\alpha_1^+ - 1}}$ and is always integrable.

This simple example shows that one can change freely the distribution of large jumps (as long as the new Lévy measure is absolutely continuous with respect to the old one) but one should be very careful with the distribution of small jumps (which is determined by the behavior of the Lévy measure near zero). This is a good property since large jumps are the ones which are important from the point of view of option pricing: they affect the tail of the return distribution and option prices in an important way. This is precisely the degree of freedom we will use in order to calibrate option prices while remaining in a class of measures equivalent to a given one.

Compound Poisson case A compound Poisson process is a pure jump Lévy process which has almost surely a finite number of jumps in every interval. This means that if two Lévy processes satisfy the conditions of mutual absolute continuity listed in proposition 1 and one of them is of compound Poisson type, the other one will also be of compound Poisson type since these processes must have the same almost sure behavior of sample functions. If the jump parts of both Lévy processes are of compound Poisson type the conditions of the proposition 1 are somewhat simplified:

Corollary 1. *Suppose that the jump part of X_t is of compound Poisson type. Then $P|_{\mathcal{F}_t}$ and $P'|_{\mathcal{F}_t}$ are mutually absolutely continuous for all t if and only if the following conditions are satisfied:*

1. $a = a'$
2. *The jump part of X'_t is of compound Poisson type and the two jump size distributions are mutually absolutely continuous.*
3. *If $a = 0$ then we must in addition have $\gamma' = \gamma$*

The Radon-Nikodym derivative is given by

$$\frac{dP'|_{\mathcal{F}_t}}{dP|_{\mathcal{F}_t}} = e^{U_t} \tag{19}$$

where U_t is a Lévy process with jump part of compound Poisson type. Its characteristic triplet is given by (15)-(17).

Proof. First of all, the condition (13) is fulfilled automatically as

$$\int_{-\infty}^{\infty} (e^{\phi(x)/2} - 1)^2 \nu(dx) \le 2 \int_{-\infty}^{\infty} (\nu(dx) + \nu'(dx)) < \infty \tag{20}$$

As can be seen from the form of its characteristic triplet (15)-(17), the Radon-Nikodym derivative process U_t also has jump part of compound Poisson type because

$$\int_{-1}^{1} \nu_U(dx) = \int_{-1}^{1} [\nu\phi^{-1}](dx) = \int_{-1\leq\phi(y)\leq 1} \nu(dy) < \infty \tag{21}$$

□

2.4 Relative entropy for Lévy processes

The notion of *relative entropy* or *Kullback-Leibler distance* is often used as measure of closeness of two equivalent probability measures. In this section we recall its definition and properties and compute the relative entropy of the measures generated by two risk neutral exp-Lévy models.

Define $(\Omega, \mathcal{F})$ as the space of real-valued cadlag functions defined on $[0, T]$. Let $\mathbb{P}$ and $\mathbb{Q}$ be two equivalent probability measures on this path space. The relative entropy of $\mathbb{Q}$ with respect to $\mathbb{P}$ is defined as

$$\mathcal{E} = \int_{\Omega} \ln(\frac{d\mathbb{Q}}{d\mathbb{P}}) d\mathbb{Q}$$

If we introduce the function $f(x) = x \ln x$, which is clearly convex, we can write the relative entropy

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

It is readily observed that the relative entropy is a convex functional of $\mathbb{Q}$. Jensen's inequality shows that it is always non-negative:

$$\mathcal{E} = E^{\mathbb{P}}[f(\frac{d\mathbb{Q}}{d\mathbb{P}}] \geq f(E^{\mathbb{P}}[\frac{d\mathbb{Q}}{d\mathbb{P}}]) = f(1) = 0$$

As the relative entropy is equal to zero when $\frac{d\mathbb{Q}}{d\mathbb{P}} = 1$ almost surely, it follows from the convexity that it is equal to zero *only if* $\frac{d\mathbb{Q}}{d\mathbb{P}} = 1$ almost surely. The following result shows that, in the case where the measures are generated by exponential Lévy models, the relative entropy can be expressed in terms of the Lévy measures:

Proposition 2. *Let $\mathbb{P}$ and $\mathbb{Q}$ be equivalent measures on $(\Omega, \mathcal{F})$ generated by exponential Lévy models with Lévy triplets (a,γ^P,ν^P) and (a,γ^Q,ν^Q). The relative entropy $\mathcal{E}(\mathbb{Q}, \mathbb{P})$ is then given by:*

$$\mathcal{E}(\mathbb{Q}|\mathbb{P}) = \frac{T}{2\sigma^2} \left\{ \gamma^Q - \gamma^P - \int_{-1}^{1} x(\nu^Q - \nu^P)(dx) \right\}^2 + \\ T \int_{-\infty}^{\infty} (\frac{d\nu^Q}{d\nu^P} \ln(\frac{d\nu^Q}{d\nu^P}) + 1 - \frac{d\nu^Q}{d\nu^P})\nu^P(dx) \tag{22}$$

If $\mathbb{P}$ and $\mathbb{Q}$ are both risk neutral measures, the relative entropy reduces to:

$$\mathcal{E}(\mathbb{Q}|\mathbb{P}) = \frac{T}{2a}\left\{\int_{-\infty}^{\infty}(e^x-1)(\nu^Q-\nu^P)(dx)\right\}^2 \\ + T\int_{-\infty}^{\infty}\left(\frac{d\nu^Q}{d\nu^P}\ln\left(\frac{d\nu^Q}{d\nu^P}\right)+1-\frac{d\nu^Q}{d\nu^P}\right)\nu^P(dx) \quad (23)$$

Proof. Consider an exponential Lévy processes defined by (7). From the bijectivity of the exponential it is clear that the filtrations generated by X_t and S_t coincide. It is therefore equivalent to compute the relative entropy of the log-price processes (which are Lévy processes). To compute the relative entropy of two Lévy processes we will use expression (14) for Radon-Nikodym derivative:

$$\mathcal{E} = \int \ln\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\frac{d\mathbb{Q}}{d\mathbb{P}}d\mathbb{P} = E^P[U_T e^{U_T}] \quad (24)$$

where (U_t) is a Lévy process with characteristic triplet given by formulae (15) - (17). Let $\phi_t(z)$ denote its characteristic function and $\psi(z)$ its characteristic exponent, that is,

$$\phi_t(z) = E^P[e^{izU_t}] = e^{t\psi(z)}$$

Then we can write:

$$E^P[U_T e^{U_T}] = -i\frac{d}{dz}\phi_T(-i) = -iTe^{T\psi(-i)}\psi'(-i) \\ = -iT\psi'(-i)E^P[e^{U_T}] = -iT\psi'(-i)$$

From the Lévy-Khinchin formula we know that

$$\psi'(z) = -a^U z + i\gamma^U + \int_{-\infty}^{\infty}(ixe^{izx} - ix1_{|x|\le 1})\nu^U(dx)$$

We can now compute the relative entropy as follows:

$$\mathcal{E} = a^U T + \gamma^U T + T\int_{-\infty}^{\infty}(xe^x - x1_{|x|\le 1})\nu^U(dx)$$
$$= \frac{T}{2}a\eta^2 + T\int(ye^y - e^y + 1)(\nu^P\phi^{-1})(dy) = \frac{T}{2}a\eta^2 + T\int\left(\frac{d\nu^Q}{d\nu^P}\ln\left(\frac{d\nu^Q}{d\nu^P}\right)+1-\frac{d\nu^Q}{d\nu^P}\right)\nu^P(dx)$$

where η is chosen such that

$$\gamma^Q - \gamma^P - \int_{-1}^{1}x(\nu^Q-\nu^P)(dx) = a\eta$$

Since we have assumed $a > 0$, we can write

$$\frac{1}{2}a\eta^2 = \frac{1}{2a}\left\{\gamma^Q - \gamma^P - \int_{-1}^{1}x(\nu^Q-\nu^P)(dx)\right\}^2$$

which leads to (22). If P and Q are martingale measures, we can express the drift γ using a and ν:

$$\frac{1}{2}a\eta^2 = \frac{1}{2a}\left\{\int_{-\infty}^{\infty}(e^x - 1)(\nu^Q - \nu^P)(dx)\right\}^2$$

Substituting the above in (22) yields (23). □

Observe that, due to time homogeneity of the processes, the relative entropy (22) or (23) is a linear function of T: the relative entropy per unit time is finite and constant. The first term in the relative entropy (22) of the two Lévy processes penalizes the difference of drifts and the second one penalizes the difference of Lévy measures.

In the risk neutral case the relative entropy only depends on the two Lévy measures ν^P, ν^Q. For a given reference measure ν^P, expression (23) viewed as a function of ν^Q defines a positive (possibly infinite) functional on the set of Lévy measures $\mathcal{L}(\mathbb{R})$:

$$\begin{aligned} H : \mathcal{L}(\mathbb{R}) \quad &\rightarrow \quad [0, \infty] \\ \nu^Q \quad &\rightarrow \quad H(\nu^Q) = \mathcal{E}(\mathbb{Q}(\nu^Q, \sigma)), \mathbb{P}(\nu^P, \sigma)) \end{aligned} \tag{25}$$

We shall call H the relative entropy functional. Its expression is given by (23). It is a positive convex functional of ν^Q, equal to zero only when $\nu^Q \equiv \nu^P$.

Compound Poisson case When the jump parts of both Lévy processes are of compound Poisson type with jump intensities λ^Q and λ^P and jump size distributions μ^Q and μ^P, the relative entropy takes the following form in the risk neutral case:

$$\begin{aligned} \frac{\mathcal{E}}{T} = \frac{\lambda^Q}{\lambda^P}\ln\frac{\lambda^Q}{\lambda^P} + \lambda^P - \lambda^Q + \frac{\lambda^Q}{\lambda^P}\int_{-\infty}^{\infty}\ln\left(\frac{\mu^Q(x)}{\mu^P(x)}\right)\mu^Q(x)dx \\ + \frac{1}{2a}\left\{\int_{-\infty}^{\infty}dx(e^x - 1)(\lambda^P\mu^P(x) - \lambda^Q\mu^Q(x))\right\}^2 \end{aligned} \tag{26}$$

2.5 Examples

Example 1: Consider two tempered stable processes that are mutually absolutely continuous and have Lévy densities given by:

$$\begin{aligned} \nu^Q(x) &= \frac{e^{(-\beta_1 - 1)x}}{x^{1+\alpha}}1_{x\geq 0} + \frac{e^{(-\beta_1+1)|x|}}{|x|^{1+\alpha}}1_{x<0} \\ \nu^P(x) &= \frac{e^{(-\beta_2 - 1)x}}{x^{1+\alpha}}1_{x\geq 0} + \frac{e^{(-\beta_2+1)|x|}}{|x|^{1+\alpha}}1_{x<0} \end{aligned}$$

with $\beta_1 > 1$ and $\beta_2 > 1$ imposed by the no-arbitrage property. The relative entropy of $\mathbb{Q}$ with respect to $\mathbb{P}$ will always be finite because we can write for the first term in (23) (we consider for definiteness the positive half-axis):

$$\int_{-\infty}^{\infty} (e^x - 1)(\nu^Q - \nu^P)(dx) = \int_{-\infty}^{\infty} dx \frac{(1-e^{-x})(e^{-\beta_1 x} - e^{-\beta_2 x})}{x^\alpha}$$

which is finite because for small x the numerator is equivalent to x^2 and for large x it decays exponentially. For the second term in (23) on the positive half-axis we have:

$$\int_{-\infty}^{\infty} (\frac{d\nu^Q}{d\nu^P} \ln(\frac{d\nu^Q}{d\nu^P}) + 1 - \frac{d\nu^Q}{d\nu^P})\nu^P(dx) \tag{27}$$

$$= \int_{-\infty}^{\infty} \frac{e^{(-\beta_2 - 1)x} - e^{(-\beta_1 - 1)x} - x(\beta_1 - \beta_2)e^{(-\beta_1 - 1)x}}{x^\alpha} \tag{28}$$

which is again finite because for small x the numerator is equivalent to x^2 and for large x we have exponential decay.

Example 2: Suppose now that in the previous example $\alpha = 1$, $\beta_1 = 2$ and $\beta_2 = 1$. In this case, although $\mathbb{Q}$ and $\mathbb{P}$ are equivalent, the relative entropy of $\mathbb{Q}$ with respect to $\mathbb{P}$ is infinite. Indeed, on the negative half-axis $\frac{d\nu^Q}{d\nu^P} = e^{|x|}$ and the criterion 13 of absolute continuity is satisfied but the $\frac{d\nu^Q}{d\nu^P} \ln\left(\frac{d\nu^Q}{d\nu^P}\right) d\nu^P = \frac{1}{|x|}$ and the second term in (23) diverges at infinity.

3 The calibration problem for exp-Lévy models

The calibration problem consists in identifying the Lévy measure ν and the volatility σ from a set of observations of call option prices. If we knew call option prices for one maturity and all strikes, we could deduce the volatility and the Lévy measure in the following way:

- Compute the risk-neutral distribution of log price from option prices using the Breeden-Litzenberger formula

$$q_T(k) = e^{-k}\{C''(k) - C'(k)\} \tag{29}$$

  where $k = \ln K$ is the log strike.

- Compute the characteristic function (2) of the stock price by taking the Fourier transform of q_T.

- Deduce σ and the Lévy measure from the characteristic function. This is particularly easy in the compound Poisson case, since the third term in the exponent in (2) is bounded. One has:

$$\sigma^2 = \lim_{u\to\infty} -\frac{2\ln\phi_T(u)}{Tu^2}$$

$$\gamma = \lim_{u\to\infty} \frac{\frac{1}{T}\ln\phi_T(u) + \frac{1}{2}\sigma^2 u^2}{iu}$$

  and the Lévy measure ν can be found by Fourier inversion.

Thus, if we knew with absolute precision a set of call option prices for all strikes and a single maturity we could deduce all parameters of our model and thus compute option prices for other maturities. In this case, option price data for any other maturity can only contradict the information we already have but cannot give us any further information. The procedure described above is however not applicable in practice for at least three different reasons. First, we do not know call prices for all strike prices but for only for a finite number of them. Actually this number may be quite small (between 10 and 40 in the empirical examples given below). Therefore the derivatives and limits in the formulae above are actually extrapolations and interpolations of the data and our inverse problem is largely under-determined. Second, even if option prices were known for all strikes and maturities, the data generating process is probably not within the exponential Lévy class due to specification error: for example, it is well known that the term structure of implied volatilities is not correctly reproduced by such models [32]. Therefore the problem (1) with equality constraints will typically have no solution: one can hope at best for a solution approximately verifying the constraints. The third difficulty is due to the presence of observational errors (or simply bid-ask spreads) in the market data. Taking derivatives of observations as in (29) can amplify these errors, rendering unstable the result of the computation. For these reasons, it is necessary to reformulate problem (1) as an *approximation* problem.

3.1 Non-linear least squares

In order to obtain a practical solution to the calibration problem, many authors have resorted to minimizing the in-sample quadratic pricing error (see for example [4, 7]):

$$(\sigma, \nu) = \arg\inf \sum_{i=1}^{N} \omega_i |C^{\sigma,\nu}(t_0, S_0, T_i, K_i) - C^*_{t_0}(T_i, K_i)|^2 \tag{30}$$

the optimization being usually done by a gradient-based method. While, contrarily to (1), one can always find some solution, the minimization functional is non-convex so a gradient descent may not succeed in locating the minimum. Given that the number of calibration constraints (option prices) is finite (and not very large), there may be many Lévy triplets which reproduce call prices with equal precision and this means the pricing error can have many local minima or, more typically, the error landscape will have flat regions in which the error has a low sensitivity to variations in model parameters (see below).

As a result the calibrated Lévy measure is very sensitive not only to the input prices but also to the numerical starting point in the minimization algorithm. Figure 1 shows an example of this instability: the two graphs represent the result of a non-linear least squares minimization where the variable is the vector of discretized values of ν on a grid. In both cases the same option prices are used, the only difference being the starting points of the optimization routines. In the first case a Merton model with intensity $\lambda = 1$ is used, in the second a

Merton model with intensity $\lambda = 5$. As can be seen in figure 1, the results of the minimization are totally different! One may think that in a parametric

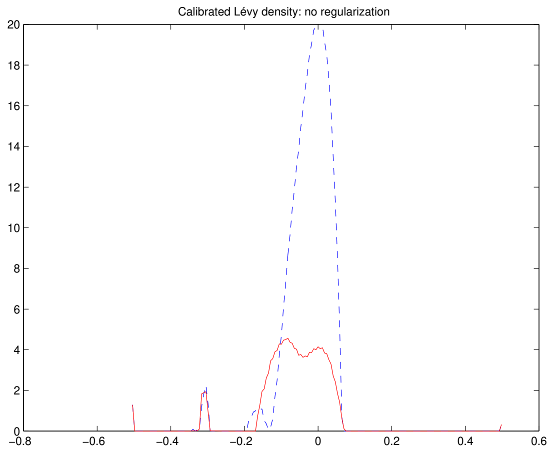

Figure 1: Lévy measure calibrated to DAX option prices, maturity 3 months via non-linear least squares method. The starting measure for both graphs is a Gaussian; the jump intensity is initialized to 1 for the red curve and to 5 for the blue one.

model with few parameters one will not encounter this problem of multiple minima since there are (many) more options than parameters. This is in fact not true, as illustrated by the following empirical example. Figure 2 represents the magnitude of the quadratic pricing error for the Merton model [27] on a data set of DAX index options, as a function of the diffusion coefficient σ and the jump intensity λ, other parameters remaining fixed. It can be observed that if one increases the jump intensity while decreasing the diffusion volatility in a suitable manner the calibration error stays approximately at the same level, leading to a flat direction in the error landscape. In fact the number of parameters is much less important from a numerical point of view than the *convexity* of the objective function to be minimized.

3.2 Regularization

The above remarks show that reformulating the calibration problem into a non-linear least squares problem does not resolve the uniqueness and stability issues: the inverse problem remains ill-posed. To obtain a unique solution in a stable manner we must introduce a regularization method [16]. One way to enforce uniqueness and stability of the solution is to add to the least-squares criterion (32) a *penalization* term:

$$(\sigma^*, \nu^*) = \arg\inf \sum_{i=1}^{N} \omega_i |C^{\sigma,\nu}(t_0, S_0, T_i, K_i) - C^*_{t_0}(T_i, K_i)|^2 + \alpha F(\mathbb{Q}, \mathbb{Q}_0) \quad (31)$$

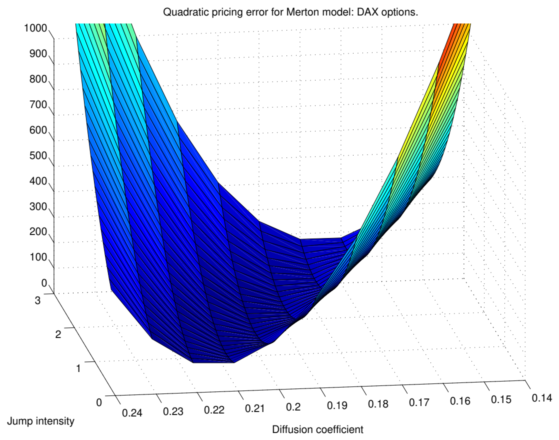

Figure 2: Quadratic pricing error as a function of model parameters, Merton model, DAX options.

where the term F, which is a measure of closeness of the model $\mathbb{Q}$ to a prior $\mathbb{Q}_0$, is chosen such that the problem (31) becomes well-posed. Problem (31) can be understood as that of finding an jump-diffusion model satisfying the conditions (1), which is closest in some sense –defined by $F(\mathbb{Q}, \mathbb{Q}_0)$– to a prior (jump-diffusion) model.

Many choices are possible for the penalization term. From the point of view of uniqueness and stability of the solution, the criterion used should be convex with respect to the parameters (here, the Lévy measure). It is this convexity which was lacking in the nonlinear least squares criterion (41).

A useful and widely used regularization criterion is provided by the *relative entropy* or *Kullback Leibler distance* $\mathcal{E}(\mathbb{Q}, \mathbb{Q}_0)$ of the the pricing measure $\mathbb{Q}$ with respect to some prior model $\mathbb{Q}_0$.

The relative entropy has several interesting properties which make it a popular choice as a regularization criterion [16]. First, as explained in section 2.4, the relative entropy plays the role of a pseudo-distance of the (risk-neutral) measure from the prior. Moreover the relative entropy becomes infinite if $\mathbb{Q}$ is not absolutely continuous with respect to the prior: using it as penalty function therefore guarantees that the solution will be a positive measure, absolutely continuous with respect to the prior. From the point of view of information theory minimizing relative entropy with respect to some prior measure corresponds to adding the least possible amount of information to the prior in order to correctly reproduce observed option prices. Finally, the relative entropy of $\mathbb{Q}$ with respect to $\mathbb{Q}_0$ is an explicitly computable functional $H(\nu)$ of the Lévy measure ν: it is given by (25). As remarked above H is a convex functional of the Lévy measure ν, with a unique minimum at $\nu = \nu_0$.

The prior probability measure with respect to which relative entropy will be calculated, may correspond for example to a jump-diffusion model estimated from historical data. In this case one can infer it from the historical data on the underlying. This is not the only possibility: the choice of the prior measure will be discussed in more detail in section 4.2.

The calibration problem now takes the form:

Calibration Problem 2. *Given a prior jump-diffusion model $\mathbb{Q}_0$ with characteristics (σ_0, ν_0) find a Lévy measure ν which minimizes*

$$\mathcal{J}(\nu) = \alpha H(\nu) + \sum_{i=1}^{N} \omega_i (C^{\nu}_{T_i}(k_i) - C^*(T_i, K_i))^2 \tag{32}$$

where $H(\nu)$ is the relative entropy of the risk neutral measure with respect to the prior, whose expression is given by (25). Here the weights ω_i are positive and sum up to one; they reflect the relative importance of reproducing different option prices precisely. For example, they may reflect the width of corresponding bid-ask intervals:

$$\omega_i = \frac{1}{|C_i^{\text{bid}} - C_i^{\text{ask}}|} \tag{33}$$

The choice of weights is addressed in more detail in section 4.1.

The functional (32) therefore consists of two parts: the relative entropy functional, which is convex in its argument ν and the quadratic pricing error which measures the precision of calibration. The coefficient α, called *regularization parameter* defines the relative importance of the two terms: it characterizes the trade-off between prior knowledge of the Lévy measure and the information contained in option prices.

In order to implement the algorithm numerically without imposing some a priori parametric form on the Lévy measure, we discretize the Lévy measure on a grid. This is done by first localizing the Lévy measure on some bounded interval $[-M, M]$ and then choosing a partition $\pi = (-M = x_1 < ... < x_N = M)$ of $[-M, M]$. Define now L_π as the set of Lévy measures with support in π:

$$L_\pi = \left\{ \sum_{x \in \pi} a(x)\delta_x, \qquad a \in (\mathbb{R}^+)^\pi \right\} \tag{34}$$

The discretized problem now becomes:

$$\inf_{\nu \in L_\pi} \mathcal{J}(\nu) \tag{35}$$

The following proposition shows that the use of entropy penalization makes our problem well posed hence numerically feasible:

Proposition 3 (Well-posedness of the regularized problem after discretization). *(i) For any partition π of $[-M, M]$, the discretized calibration problem (32)-(35) admits a solution:*

$$\exists \nu^\pi \in L_\pi, J(\nu^\pi) = \inf_{\nu \in L_\pi} \mathcal{J}(\nu) \tag{36}$$

If in addition the volatility coefficient σ is non-zero then for α large enough the solution is unique.

(ii) Every solution ν^π of the regularized problem depends continuously on the vector of input prices $(C^*(T_i, K_i), i = 1..n)$ and for a suitable choice of α converges to a minimum entropy least squares solution when the error level on input prices tends to zero.

Proof. See appendices B and C. □

If α is large enough the convexity properties of the entropy functional stabilize the solution of problem (32). When $\alpha \to 0$, we recover the non-linear least squares criterion (41). Therefore the correct choice of α is important: it cannot be fixed in advance but its 'optimal' value depends on the data at hand and the level of error δ (see section 4.3).

3.3 Relation with previous literature

3.3.1 Relation with minimal entropy martingale measures

The concept of relative entropy has been used in several contexts as a criterion for choosing pricing measures [2, 15, 17, 19, 21, 18, 28]. We briefly recall them here in relation to the present work.

In the absence of calibration constraints, the problem studied above reduces to that of identifying the equivalent martingale measure with minimal relative entropy with respect to a prior model. This problem has been widely studied and it is known that this unique pricing measure (minimal entropy martingale measure) defines the "least favorable market completion" in the sense that it minimizes the exponential utility of the optimal trading strategy [15, 17, 18]. It satisfies:

$$\mathbb{Q} = \arg\min_Q \max_X \{E^P(u(e + X - E^Q(X)))\}$$

where the min is taken over all equivalent martingale measures, the maximum is taken over all $\mathcal{F}_t$-measurable random variables, P is the historical measure and e the initial capital. $\max_X\{E^P(u(e + X - E^Q(X)))\}$ is the maximum expected (exponential) utility that can be obtained by trading in derivatives and the underlying without constraints in a market where the prices are determined by $\mathbb{Q}$. Although we only consider here the class of measures corresponding to Lévy processes, if the prior measure is a Lévy process then the MEMM is known to define again a Lévy process [28]. However the notion of MEMM does not take into account the information obtained from observed option prices.

To take into account the prices of derivative products traded in the market, Kallsen [21] introduced the notion of consistent pricing measure, that is, a measure that correctly reproduces the market-quoted prices for a given number of derivative products. He studies the relation of the minimal entropy consistent martingale measure (the martingale measure that minimizes the relative entropy distance to a given prior and respects a given number of market prices) to exponential hedging. He finds that this MECMM defines the "least favorable consistent market completion" in the sense that it minimizes the exponential utility of the optimal trading strategy over all consistent martingale measures (see also [15]). It satisfies:

$$Q = \arg\min_{Q} \max_{X} \{E^P(u(e + X - E^Q(X)))\}$$

where the min is taken over all consistent equivalent martingale measures, the max is taken over all $\mathcal{F}_T$-measurable random variables, P is the prior/historical measure and e the initial capital.

The minimal entropy measure studied in this article is not equivalent to the MECMM studied by Kallsen because we impose an additional restriction that the calibrated measure should stay in the class of measures corresponding to Lévy processes. It can be shown that the two measures only coincide in the case where there is no calibration constraints. However, in the case where calibration constraints are present our measure can be seen as an approximation of the MECMM which stays in the class of Lévy processes. The usefulness of this approximation is clear: whereas the MECMM is an abstract notion for which one can at most assert existence and uniqueness, the one studied here is actually computable (see below) and can easily be used directly for pricing purposes. Therefore our framework can be regarded as a *computable* approximation of Kallsen's minimal entropy constrained martingale measure.

3.3.2 Relation with calibration algorithms based on relative entropy minimization

In a series of papers [2, 3], Avellaneda and collaborators have proposed a non-parametric method based on relative entropy minimization for calibrating a pricing measure. In [2] the calibration problem is formulated as one of finding a pricing measure which minimizes relative entropy with respect to a prior given calibration constraints:

Calibration Problem 3.

$$\mathbb{Q} = \arg\min_{Q \sim \mathbb{Q}_0} \mathcal{E}(Q, \mathbb{Q}_0) \quad \text{under} \quad E^Q(S(T_i) - K_i)^+ = C_t^*(T_i, K_i), \quad i = 1 \ldots n \tag{37}$$

where minimization is performed over all (not necessarily "risk neutral") probability measures Q equivalent to $\mathbb{Q}_0$. Problem (37) is still ill-posed since the equality constraints may be impossible to verify simultaneously due to model mis-specification: a solution may not exist. However, it is not necessary to consider equality constraints like those in (37) since the market option prices are not exact but always quoted as bid-ask intervals. In a subsequent work, Avellaneda *et al* [3] consider a regularized version of problem (37) with quadratic penalization of constraints.

$$\mathbb{Q} = \arg\min_{Q \sim Q_0} \mathcal{E}(Q, \mathbb{Q}_0) + \sum_{i=1}^{n} |C^*(T_i, K_i) - E^Q(S(T_i) - K_i)^+|^2 \tag{38}$$

In both cases the state space is discretized and the problem solved by a dual method: the result is a calibrated (but not necessarily "risk neutral") probability distribution on a finite set of paths, generated from the prior Q_0.

Although our formulation (32) looks quite similar to (38), there are several important differences.

First, note that the numerical solution of our problem (32) is done via discretization of the *parameter* space not the state space Ω: the solution of (32) corresponds to a well defined continuous time process. By contrast in [3], the discretization is applied to the state space: in fact the Weighted Monte Carlo produces a measure $\mathbb{Q}$ which does not correspond to a continuous time process but is defined as a measure on a finite set of trajectories, namely the support of the empirical distribution generated through Monte Carlo simulation from the prior $\mathbb{Q}_0$. The continuous limit is very subtle and not easy to describe[2].

Second, while the minimization in (38) is performed over all probability measures equivalent to the prior (the optimization variables are the probabilities themselves), in our case the minimization is performed over equivalent measures corresponding to jump-diffusion (exp-Lévy) models, parametrized by their Lévy measure ν. While restricting the class of models, this approach has an advantage: it guarantees that we remain in the class of risk neutral models, which is not the case in [3].

Third, while in [3] the optimization variables is the (discretized) probability measure Q itself, in our case the optimization variable is the Lévy measure ν. As a consequence, while the Weighted Monte Carlo method yields a set of weights on trajectories, which is then used to price other options by Monte Carlo, our method yields a local description of the process (i.e. its infinitesimal generator) through knowledge of ν. In particular in order to price options one can use either Monte Carlo methods or solve the associated partial integro-differential equation, which may be preferable for American or barrier options.

Finally, even in the case Monte Carlo methods to price other options once the model is calibrated, note that in the Weighted Monte Carlo method pricing is done using the original sample paths, simulated from the prior model. Our approach has the advantage that we do not depend on the original set of paths in order to perform the Monte Carlo. Indeed, the posterior (calibrated) measure may be quite different from the prior rendering many of the initial paths useless for computing expectations under the calibrated measure. Knowing the Lévy measure ν allows to generate new paths under $\mathbb{Q}$.

[2] We thank Patrick Cattiaux for discussions on this point.

4 Numerical implementation

As explained in section 3, we tackle the ill-posedness of the initial calibration problem by transforming it into an optimization problem:

$$\begin{aligned} \nu^* &= \arg \inf_{\nu \in \mathcal{L}(\mathbb{R})} \mathcal{J}(\nu) \\ \mathcal{J}(\nu) &= \alpha H(\nu) + \sum_{i=1}^{N} \omega_i |C^{\sigma,\nu}(t_0, S_0, T_i, K_i) - C^*_{t_0}(T_i, K_i)|^2 \end{aligned} \tag{39}$$

We now describe a numerical algorithm for solving the optimization problem (39). There are four main steps in the numerical solution:

- Choice of the weights assigned to each option in the objective function.
- Choice of the prior measure $\mathbb{Q}_0$ from the data.
- Choice of the regularization parameter α.
- Solution of the optimization problem (39) for given α and $\mathbb{Q}_0$.

We shall describe each of these steps in detail below. This sequence of steps can be repeated a few times in order to minimize the influence of the choice of the prior.

4.1 The choice of weights in the minimization functional

The relative weights ω_i of option prices in the minimization functional should reflect our confidence in individual data points which is determined by the liquidity of a given option. This can be assessed from the bid-ask spreads, but the bid and ask prices are not always available from option price data bases. On the other hand, it is known that at least for the options that are not too far from the money, the bid-ask spreads is of order of tens of basis points ($< 1\%$). This means that in order to have errors proportional to bid-ask spreads, one must minimize the differences of implied volatilities and not those of the option prices. However, this method involves many computational difficulties (numerical inversion of the Black-Scholes formula at each minimization step). A feasible solution to this problem is to minimize the square differences of option prices weighted by the Black Scholes "vegas" evaluated at the implied volatilities of the market option prices.

$$\sum_{i=1}^{N} (I(C^{\nu}_{T_i}(k_i)) - I_i)^2 \approx \sum_{i=1}^{N} \frac{\partial I}{\partial C}(I_i) |C^{\nu}_{T_i}(k_i) - C^*_i|^2 = \sum_{i=1}^{N} \frac{(C^{\nu}_{T_i}(k_i) - C_i)^2}{\mathrm{Vega}^2(I_i)} \tag{40}$$

where $I(.)$ denotes the Black Scholes implied volatility as a function of option price and I_i denotes the market implied volatilities.

4.2 Determination of the prior

From a "Bayesian" perspective, one would expect the user to specify a prior: in this case, the user would have to specify a Lévy measure ν_0 and a diffusion coefficient σ_0. For example, these could be the result of the statistical estimation of a jump diffusion model for the time series of asset returns. However, typically the user may not have such detailed views and it is important to have a procedure to generate a reference measure $\mathbb{Q}_0$ automatically from options data. To do this we use an auxiliary jump-diffusion model (e.g. Merton model) described by the volatility parameter σ_0 and a few other variables (denoted by θ) parametrizing the Lévy measure: $\nu_0 = \nu_0(\theta)$. This model is then calibrated to data using the standard least squares procedure (32):

$$\begin{aligned}(\sigma_0, \nu_0) &= \arg\inf_{\sigma,\theta} \epsilon(\sigma, \nu(\theta)) \\ \epsilon(\sigma, \nu(\theta)) &= \sum_{i=1}^{N} \omega_i |C^{\sigma,\nu(\theta)}(t_0, S_0, T_i, K_i) - C^*_{t_0}(T_i, K_i)|^2 \end{aligned} \tag{41}$$

Since the objective function is not convex, a simple gradient procedure may not give the global minimum. However, as we will see, the solution (σ_0, ν_0) will be iteratively improved at later stages and should only be viewed as a way to regularize the optimization problem (39) so the minimization procedure at this stage need not be very precise.

4.3 Determination of the regularization parameter

As remarked above, the regularization parameter α determines the tradeoff between the accuracy of calibration and the numerical stability of the results with respect to the input option prices. It is therfore plausible that the right value of α should depend on the data at hand and should not be determined a priori.

One way to achieve this tradeoff is by using the Morozov discrepancy principle [16]. First, we minimize the quadratic pricing error (30). The value of $\epsilon(\sigma_0, \nu_0)$ of this optimization problem can now be interpreted as a measure of "model error": if $\epsilon(\sigma_0, \nu_0) = 0$ then it means that perfect calibration is achieved by the prior but typically $\epsilon(\sigma_0, \nu_0) = \epsilon_0 > 0$ where ϵ_0 represents the the 'distance' of market prices to model prices i.e. it gives an a priori level of quadratic pricing error that one cannot really hope to improve upon while staying in the same class of models. Note that here we only need to find the minimal value of (30) and not to locate its minimum so a rough estimate is sufficient and the presence of "flat" directions is not a problem.

Now let $(\sigma_\alpha, \nu_\alpha)$ be the solution of (39) for a given regularization parameter $\alpha > 0$. Then the *a posteriori* quadratic pricing error is given by $\epsilon(\sigma_\alpha, \nu_\alpha)$, which one expects to be a bit larger than ϵ_0 since by adding the entropy term we have sacrificed some precision in order to gain in stability. The Morozov discrepancy principle consists in minimizing this loss of precision through regularization by choosing α such that

$$\epsilon(\sigma_\alpha, \nu_\alpha) \simeq \epsilon_0 \tag{42}$$

In pratice we fix some $\delta > 1, \delta \simeq 1$ (for example $\delta = 1.1$) and numerically solve

$$\epsilon(\sigma_\alpha, \nu_\alpha) = \delta\epsilon_0 \tag{43}$$

The left hand side is a differentiable function of α so the solution can be obtained with a small number of iterations for example by Newton's (or a dichotomy) method with a few iterations.

4.4 Computation of the gradient

In order to minimize the functional (39) using a BFGS gradient descent method, the essential step is the computation of the gradient. We represent the Lévy measure ν by discretizing it on a grid $(x_i, i = 1..N)$ where $x_i = x_0 + i\Delta x$. The grid must be uniform in order to use the FFT algorithm for option pricing. This means that we effectively allow a fixed (but large) number of jump sizes and calibrate the intensities of these jumps. The Lévy process is then represented as a weighted sum of independent standard Poisson processes with different intensities, which is none other than the discretization of the Lévy Khinchin representation (2).

In order to use the BFGS gradient descent method to find the minimum, we need to compute the gradient of the functional (39) with respect to the Lévy measure ν. If we were to compute this gradient numerically, the complexity would increase by a factor equal to the number of grid points. A crucial point of the method is that we are able to compute the gradient of the option prices with only a two-fold increase of complexity compared to computing prices alone. Due to this optimization, the execution time of the program changes on average from several hours to about a minute on a standard PC.

We now compute the variational derivative of the option price. Here for the sake of simplicity all the computations are carried out in the continuous case. In the discretized case the idea is the same, but the Fréchet derivative is replaced by the usual gradient and all the formulae become more cumbersome.

The functional which maps Lévy measure into option price is defined by formulae (2) and(62). To show that all functions that we are working with are, in addition to their other arguments, functionals of the Lévy measure, we will write it as a second argument in square brackets ($z_T(k)[\nu]$). Let us take an admissible test function h and compute the directional derivative of $z_T(k)[\nu]$ in the direction h. By definition

$$D_h z_T(k)[\nu] = \frac{\partial}{\partial\varepsilon}\{z_T(k)[\nu + \varepsilon h]\}|_{\varepsilon=0} \tag{44}$$

We then obtain under sufficient integrability conditions on the stock price process, combining the formulae (2)-(62) and performing the differentiation with respect to ε that the directional derivative $D_h z_T(k)$ of the option price with respect to the Lévy density is given by:

$$D_h z_T(k)[\nu] = \frac{1}{2\pi}\int_{-\infty}^{\infty} dv e^{-ivk-rT}\frac{Te^{T\psi(v-i)}}{iv(1+iv)} \int_{-\infty}^{\infty} dx h(x)\{e^{ivx}-1-ive^x+iv\} \quad (45)$$

By interchanging the two integrals, we can compute, again under sufficient integrability conditions, the Fréchet derivative Dz_T of the option price :

$$Dz_T(k)[\nu] = \frac{1}{2\pi}\int_{-\infty}^{\infty} dv e^{-ivk-rT}\frac{Te^{T\psi(v-i)}}{iv(1+iv)}\{e^{ivx}-1-ive^x+iv\} \quad (46)$$

By rearranging terms and separating integrals we have:

$$\begin{aligned} Dz_T(k)[\nu] = \frac{T}{2\pi}\int_{-\infty}^{\infty} dv e^{-iv(k+x)}\frac{e^{-rT}e^{T\psi(v-i)}-e^{ivrT}}{iv(1+iv)} - \\ \frac{T}{2\pi}\int_{-\infty}^{\infty} dv e^{-ivk}\frac{e^{-rT}\exp(T\psi(v-i))-e^{ivrT}}{iv(1+iv)} + \\ \frac{T}{2\pi}\int_{-\infty}^{\infty} dv e^{-ivk}\frac{e^{-ivx+ivrT}-e^{ivrT}}{iv(1+iv)} - \\ \frac{T(e^x-1)}{2\pi}\int_{-\infty}^{\infty} dv e^{-ivk-rT}\frac{e^{T\psi(v-i)}}{1+iv} \end{aligned} \quad (47)$$

Here the first two terms may be expressed in terms of the option price function, the third one does not depend on the Lévy measure and can be computed analytically and the last one is a product of a simple function of x and some auxiliary function which does not depend on x (and therefore must be computed only once for each gradient evaluation). Finally we obtain:

$$\begin{aligned} Dz_T(k)[\nu] = Tz_T(k+x) - Tz_T(k) + T(1-e^{k+x-rT})^+ - T(1-e^{k-rT})^+ - \\ \frac{T(e^x-1)}{2\pi}\int_{-\infty}^{\infty} dv e^{-ivk-rT}\frac{\exp(T\psi(v-i))}{1+iv} = \\ T(C_T(k+x)-C_T(k)) - (e^x-1)\frac{T}{2\pi}\int_{-\infty}^{\infty} dv e^{-ivk-rT}\frac{\exp(T\psi(v-i))}{1+iv} \end{aligned} \quad (48)$$

Fortunately, this expression may be represented in terms of the option price and one auxiliary function. Since we are using FFT to compute option prices for the whole price sheet, we already know these prices for the whole range of strikes. As the auxiliary function will also be computed using the FFT algorithm, the computational time will only increase by a factor of two.

4.5 The algorithm

Here is the final numerical algorithm as implemented in the numerical examples given below.

1. Calibrate an auxiliary jump-diffusion model (Merton model) to obtain an estimate of volatility σ_0 and a candidate for the prior Lévy measure ν_0.

2. Estimation of the "noise level" ϵ_0:

Here we distinguish two cases, depending on the data available as input.

• If bid prices and ask prices are available for each calibration constraint, then the a priori error level can be computed as:

$$\epsilon_0^2 = \sum_{i=1}^{N} \omega_i |C_i^{bid} - C_i^{ask}|^2 \tag{49}$$

• If confidence intervals/ bid-ask intervals are not available, the a priori error level ϵ_0 must be estimated from the data themselves. We have used the following method: fix $\sigma = \sigma_0$ and run algorithm with $\alpha \simeq 0$ using gradient descent method with low precision to get estimate of "distance to model"

$$\epsilon_0^2 = \inf_{\nu} \sum_{i=1}^{N} \omega_i |C_i^{\sigma_0,\nu} - C_i^*|^2 \tag{50}$$

3. Use a posteriori method described in 4.3 to compute optimal regularization parameter α^* acheiving tradeoff between precision and stability:

$$\varepsilon(\alpha^*) = \sum_{i=1}^{N} \omega_i |C_i^{\sigma,\nu} - C_i^*|^2 \simeq \epsilon_0^2 \tag{51}$$

The optimal α^* is found by running the gradient descent method (BFGS) several times with low precision.

4. Solve variational problem for $\mathcal{J}(\nu)$ with α^* by gradient-based method (BFGS) with high precision using either a user-specified prior or result of 1) as prior.

This procedure can be iterated a few times, using as auxiliar model in step 1) the calibrated model in step 4) in order to minimize the influence of the prior on the result. In practice (see the numerical tests in the next section) we have found that a single iteration already gives satisfactory results.

5 Numerical tests

In order to verify the accuracy and numerical stability of our algorithm, we have first proceeded to test it on simulated data sets of option prices generated using a jump diffusion model. This allows us to explore the magnitude of finite sample effects and to assess the importance of the two different stages of the calibration procedure described in section 4. In the first series of tests, option prices were generated using Kou's jump diffusion model [23] with a diffusion part $\sigma_0 = 10\%$ and a Lévy density:

$$\nu(x) = \lambda[1_{x>0} p \alpha_1 e^{-\alpha_2 x} + (1-p)\alpha_2 e^{-\alpha_2 x} 1_{x<0}] \tag{52}$$

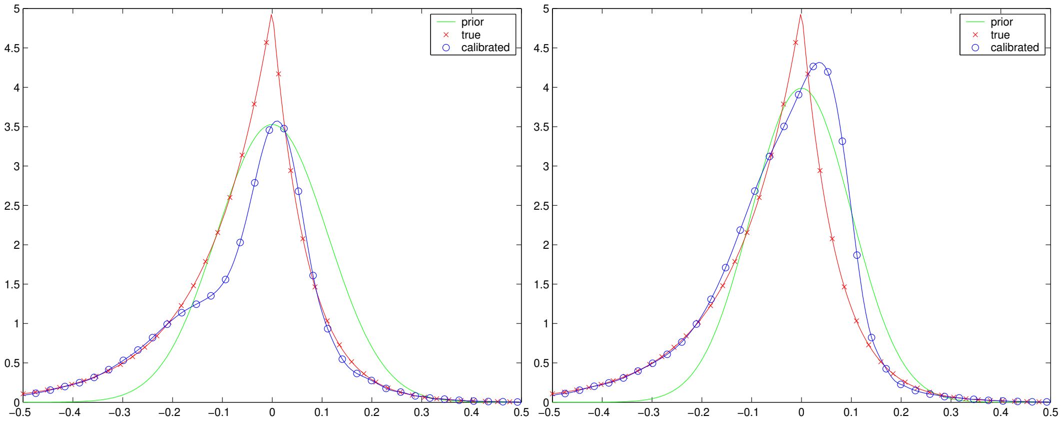

Figure 3: Lévy measure calibrated to option prices simulated from Kou's jump diffusion model with $\sigma_0 = 10\%$. Left: σ has been calibrated in a separate step ($\sigma = 10.5\%$). Right: σ was fixed to $9.5\% < \sigma_0$.

In the tests we have taken an asymmetric density with the left tail heavier than the right one ($\alpha_1 = 1/0.07$ and $\alpha_2 = 1/0.13$). The intensity was taken to be $\lambda = 1$ and the last constant p was chosen such that the density is continuous at $x = 0$. The option prices were computed using the Fourier transform method described in the appendix. The maturity of the options was 5 weeks and we used 21 equidistant strikes ranging from 6 to 14 (the spot being at 10). In order to capture tail behavior it is important to have strikes quite far in and out of the money. As the prior model we use Merton's jump diffusion model. In this model the jump part of the log price is a compound Poisson process and the jump sizes are normally distributed with mean zero:

$$X_t \quad = \quad bt + \sigma W_t + \sum_{i=1}^{N_t^\lambda} Y_i \qquad Y_i \sim N(0, \gamma^2) \text{ IID} \tag{53}$$

In Merton's model the price of a call option can be expanded as a weighted superposition of Black Scholes prices with weights exponentially converging to zero. This series expansion allows fast computation of call prices which is necessary for the first step of the algorithm described in section 4.

After generating sets of call option prices from Kou's model using the FFT method desribed in the appendix, the algorithm described in section 4 was applied to the option prices obtained. Figure 3 compares the non-parametric reconstruction of the Lévy density to the true Lévy density which, in this case, is known to be (52). As observed in figure 4, the accuracy of calibration at the level of option prices and/or implied volatilities is satisfying with only 21 options. Comparing the jump size densities obtained with the true one, we observe that we retrieve successfully the main features of the true density with our non-parametric approach. The only region in which we observe a detectable

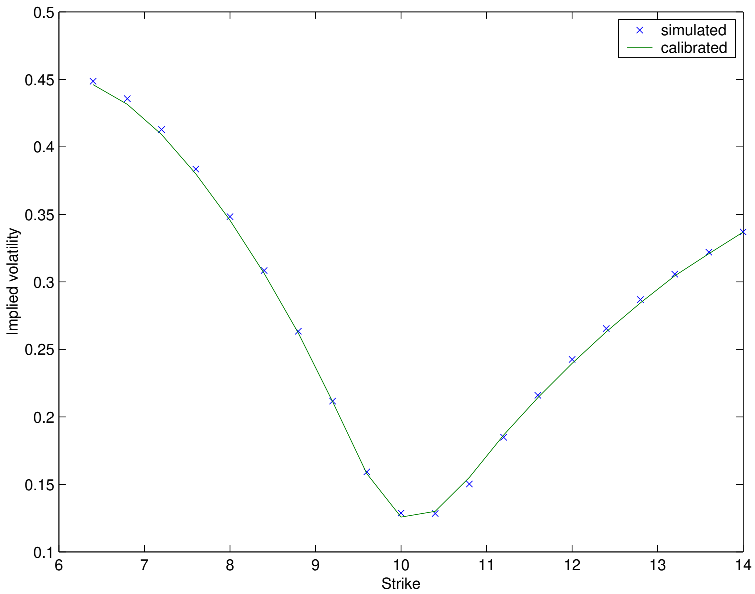

Figure 4: Calibrated vs simulated (true) implied volatilities corresponding to figure3 for Kou model [23].

error is near zero: very small jumps have a small impact on option prices. In fact, the gradient of our calibration criterion (computed in section 4.4) vanishes at zero which means that the algorithm does not modify the Lévy density in this region: the intensity of small jumps can not be retrieved accurately. The redundancy of small jumps and diffusion component is well known in the context of statistical estimation on time series [8, 26]. Here we retrieve another version of this redundancy in a context of calibration to a cross sectional data set of options.

Comparing the left and right graphs in figure 3 further illustrates the redundancy of small jumps and diffusion: the two graphs were calibrated to the same prices and only differ in the diffusion coefficients. Comparing the two graphs shows that the algorithm compensates the error in the diffusion coefficient by adding jumps to the Lévy density such that, overall, the accuracy of calibration is maintained (the standard deviation is 0.2%).

The stability of the algorithm with respect to initial conditions can be examined by perturbating the starting point of the optimization routine and examining the effect on the output. As illustrated in figure 5, the entropy penalty removes the sensitivity observed in the non-linear least squares algorithm (see figure 1 and section 3.1). The only minor difference between the two calibrated measures is observed in the neighborhood of zero i.e. the region which, as remarked above, has little influence on option prices.

In a second series of tests we examine how our method performs when applied

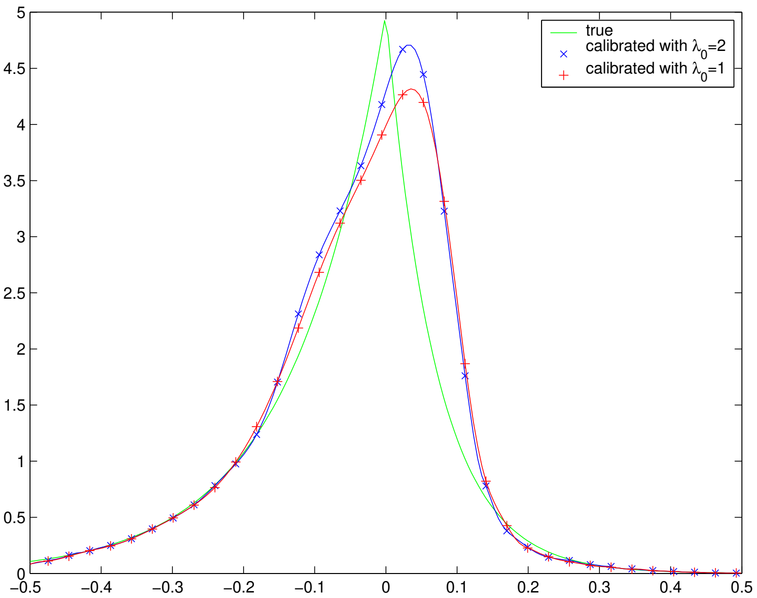

Figure 5: Levy densities calibrated to option prices generated from Kou model, using two different initial measures with intensities $\lambda = 1$ and $\lambda = 2$.

to option prices generated by an infinite activity process such as the variance gamma model. We assume that the user, ignoring that the data generating process has infinite activity, chooses a (misspecified) prior which has a finite jump intensity (e.g. the Merton model).

Option prices for 30 strike values were generated using the variance gamma model [24] with no diffusion component ($\sigma_0 = 0$) and the calibration algorithm was applied using as prior a Merton jump-diffusion model. Figure 6 shows that even though the prior is misspecified, the result reproduced the implied volatilities with good precision (the error is less than 0.5% in implied volatility units). The calibrated value of the diffusion coefficient of $\sigma = 7.5\%$, while the Lévy density has been truncated near zero to a finite value (figure 7 left): the algorithm has substituted a non-zero diffusion part for the small jumps which are the origin of infinite activity. Figure 7 further compares the Lévy measures obtained when fixing σ to two different values: we observe that a smaller value of the volatility parameter leads to a greater intensity of small jumps.

Here we observe once again the redundancy of volatility and small jumps, this time in an infinite-activity context. More precisely this example shows that call option prices generated from an infinite activity jump-diffusion model can be reproduced with arbitrary precision using a compound Poisson model with finite jump intensity. This leads us to conclude that for a finite (but realistic) number of observations, infinite activity models like variance gamma are hard to distinguish from finite activity compound Poisson models on the basis of option

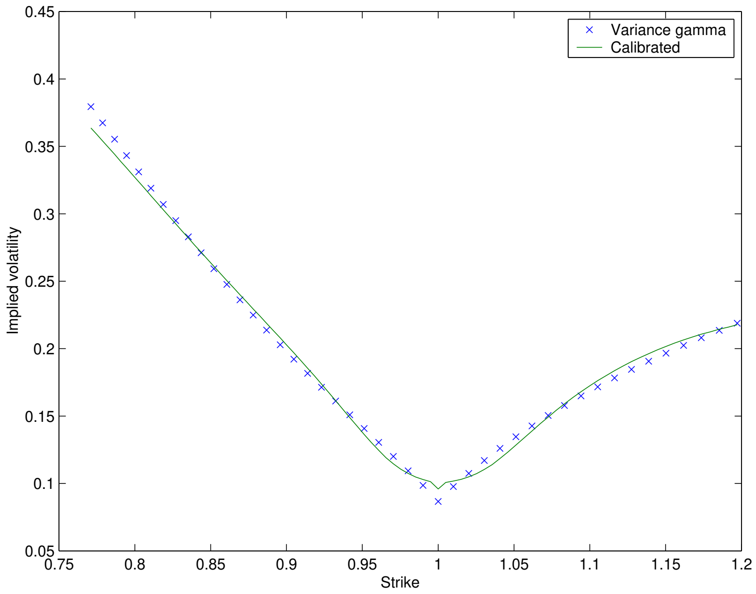

Figure 6: Implied volatility smile for variance gamma model with $\sigma_0 = 0$ compared with smile generated from the calibrated Lévy measure. Calibration yields $\sigma = 7.5\%$

prices.

6 Empirical results

To illustrate our calibration method we have applied it to a data set of daily series of prices and implied volatilities for options on the DAX (German index) from 1999 to 2001. A detailed description of data formats and filtering procedures can be found in [11]. Some of the results obtained on this data set are described below.

6.1 Empirical properties of the Lévy density

Figure 8 illustrates the typical shape of a risk neutral Lévy density obtained from our data set: it is peaked at zero and highly skewed towards negative values.

The effect of including the entropy penalty can be assessed by comparing the results obtained when changing the prior and/or the initialization in the algorithm. Figure 9 compares the Lévy measures obtained with different priors: in this case the jump intensity of the prior (a Merton model) was shifted from $\lambda = 1$ to $\lambda = 5$. Compared to the high sensitivity observed in the nonlinear

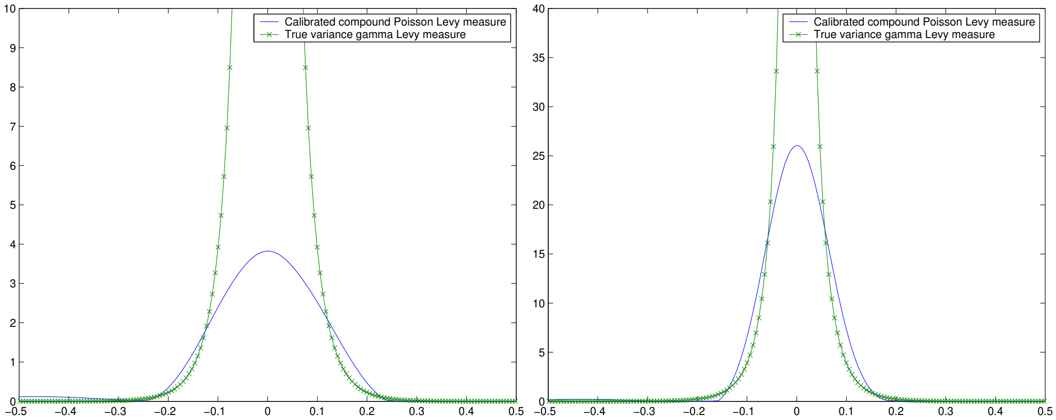

Figure 7: Lévy measure calibrated to variance gamma option prices with $\sigma = 0$ using a compound Poisson prior with $\sigma = 10\%$ (left) and $\sigma = 7.5\%$ (right). Increasing the diffusion coefficient decreases the intensity of small jumps in the calibrated measure.

least squares algorithm (figure 1), we observe that adding the entropic penalty term has stabilized our algorithm.

The logarithmic scale in figure 9 allows the tails to be seen more clearly. Recall that the prior density is gaussian, which shows up as a symmetric parabola on log scales. It is readily seen that the Lévy measures obtained are far from being symmetric: the distribution of jump sizes is highly skewed towards negative values. Figure 13 shows the same result across calendar time, showing that this asymmetry persists across time. This effect also depends on the maturity of options in question: for longer maturities (see 14) the support of the Lévy measure extends further to the left.

The *area* under the curves shown here is to be interpreted as the (risk neutral) jump intensity. While the shape of the curve does vary slightly across calendar time, the intensity stays surprisingly stable: its numerical value is empirically found to be $\lambda \simeq 1$, which means around one jump a year. Of course note that this is the *risk neutral* intensity: jump intensities are *not* invariant under equivalent change of measures. Moreover this illustrates that a small intensity of jumps λ can be sufficient for explaining the shape of the implied volatility skew for small maturities. Therefore the motivation of infinite activity processes does not seem clear to us, at least from the viewpoint of option pricing.

6.2 Testing time homogeneity

While the literature on jump processes in finance has focused on time homogeneous (Lévy) models, practitioners have tended to use time dependent jump or volatility parameters. Here we can investigate time homogeneity in a nonparametric way by separately calibrating the Lévy measure to various option

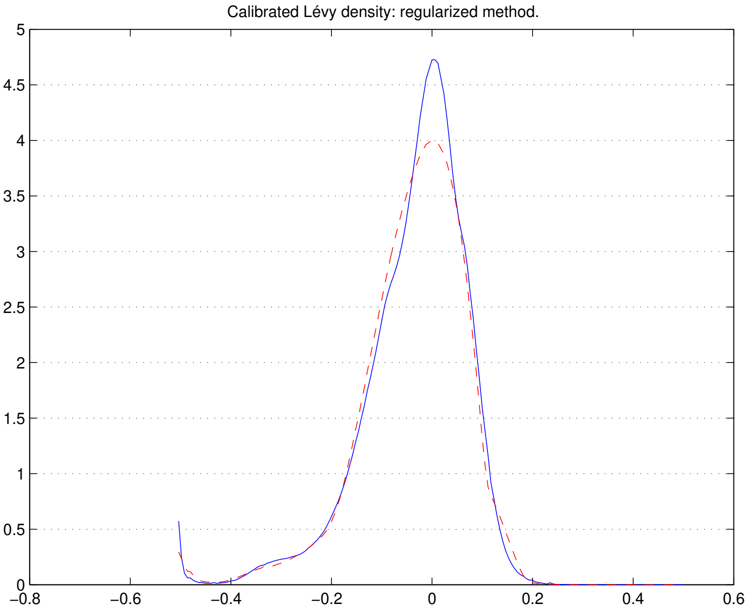

Figure 8: Lévy density calibrated to DAX option prices, maturity 3 months.

maturities. Figure 10 shows Lévy measures obtained by running the algorithm separately for options of different maturity. The null hypothesis of time homogeneity would imply that all the curves are the same, which is apparently not the case here. However computing the areas under the curves yields similar jump intensities across maturities: this result can be interpreted by saying that the risk neutral jump intensity is relatively stable through time while the shape of the (normalized) jump size density can actually change. Of course, this is a more complicated form of time dependence than simply having a time dependent intensity.

These results can be further used to investigate what form of time dependence is appropriate to introduce in order to capture the empirically observed term structure of implied volatilities. Whether introducing such time dependence in the jump density is an appropriate way to extend such models is not obvious to us.

7 Conclusion

We have proposed a non-parametric method for identifying, in a numerically stable fashion, a risk neutral jump-diffusion model consistent with market prices of options and equivalent to a prior model. We have also presented a stable computational implementation and tested its performance on simulated and empirical data. Theoretically our method can be seen as a computable approximation

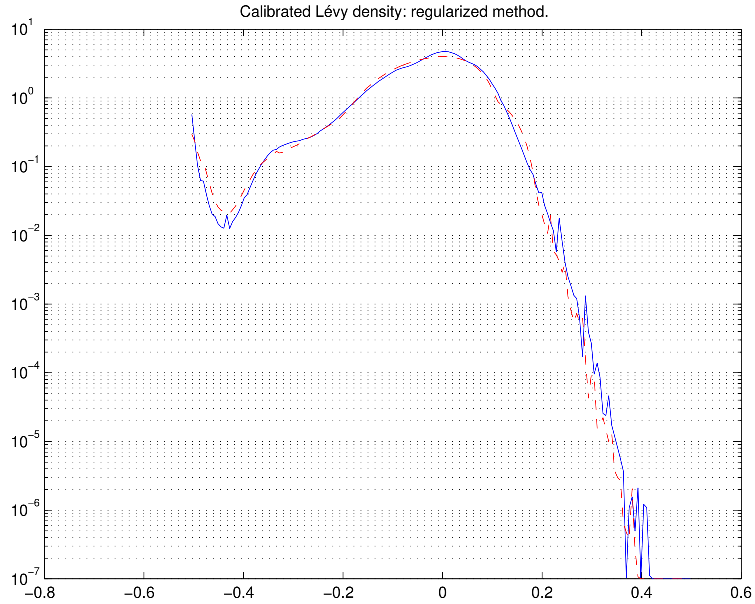

Figure 9: Logarithm of Lévy density calibrated to DAX option prices, maturity 3 months. Logarithmic scale.

to the notions of minimal entropy martingale measures, made consistent with observed market prices of options. Computationally, it is a stable alternative to current least squares calibration methods for jump-diffusion models. The jump part is retrieved in a non-parametric fashion: we do not assume shape restrictions on the Lévy measure. Finally, our approach allows to conciliate the idea of calibration by relative entropy minimization [2] with the notion of risk neutral valuation.

Our method can complement in various ways the existing literature on parametric jump-diffusion models in option pricing. First, using a non-parametric calibration is not necessarily incompatible with using a parametric model for pricing. Our method can be used as a specification test for choosing the correct parametric class of jump diffusion models. Second, we provide a computational approach for estimating risk-neutral jump processes from options data which can be potentially applied to other models where jump processes have to be deduced from observation of contingent claims: credit risk models are typically such examples. Third, separate calibration of the jump density to various option maturities can be used to investigate time inhomogeneity in a non-parametric way. Finally, our approach can be extended to other inverse problems in which an unknown jump process has to be identified, such as calibration problems for stochastic volatility models with jumps [5, 7]. We intend to pursue these issues in our future research.

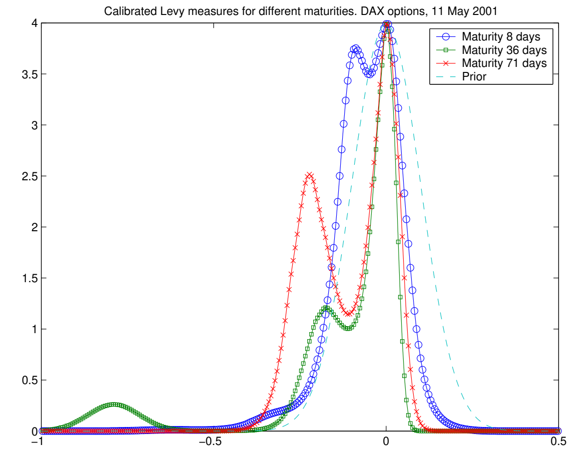

Figure 10: Lévy measures calibrated to DAX options, all maturities. Each curve corresponds to a different maturity.

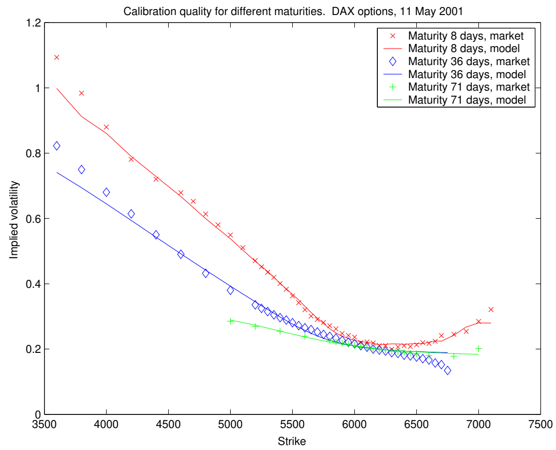

Figure 11: Calibration quality for different maturities: market implied volatilities for DAX options against model implied volatilities. Each maturity has been calibrated separately.

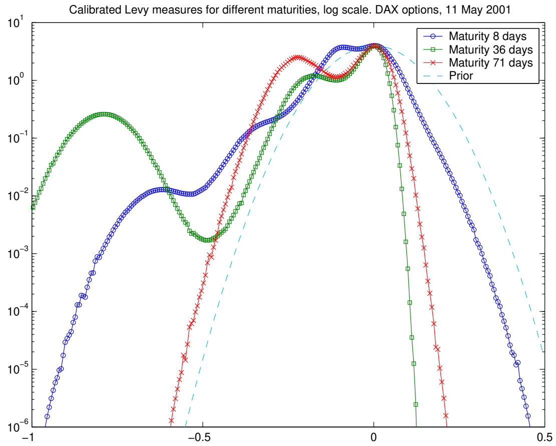

Figure 12: Lévy measures calibrated to DAX options, logarithmic scale.

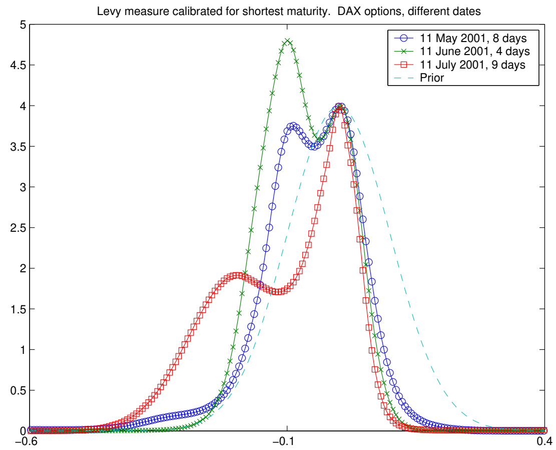

Figure 13: Results of calibration at different dates for shortest maturity. DAX index options.

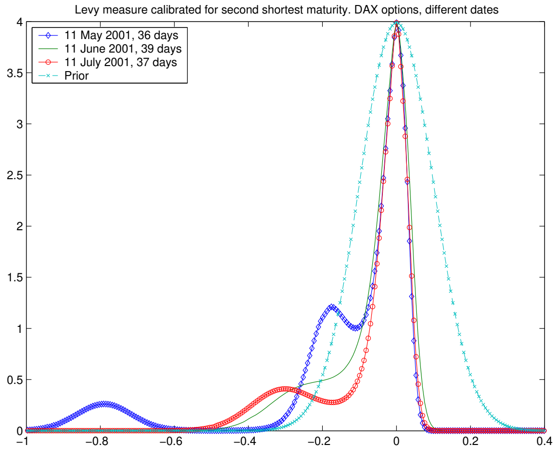

Figure 14: Results of calibration at different dates for second shortest maturity. DAX index options.

References

[1] Ait-Sahalia, Y. & Lo, A. (1998): Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices, *Journal of Finance*, 53, 499-547.

[2] Avellaneda, M. (1998) Minimum entropy calibration of asset pricing models, *International Journal of theoretical and applied finance*, **1** (4), 447-472.

[3] Avellaneda M., Buff R., Friedman C., Grandchamp N., Kruk L., Newman J. (2001): Weighted Monte Carlo: a new technique for calibrating asset pricing models, International Journal of Theoretical and Applied Finance Vol. 4, No. 1 (2001) 91-119.

[4] Andersen, L. & Andreasen, J. (2000): Jump diffusion models: volatility smile fitting and numerical methods for pricing. *Review of Derivatives Research*, **4**, 231–262.

[5] Barndorff-Nielsen, O., Mikosch T. & Resnick S. (Eds.) (2001) *Lévy processes-theory and applications*, Boston: Birkhauser.

[6] Barndorff-Nielsen,O.(1998) Processes of normal inverse Gaussian type, *Finance and Stochastics*, 2, 41-68.

[7] Bates, D.S. (1996) Jumps and stochastic volatility: the exchange rate processes implicit in Deutschemark options, *Review of financial studies*, *9*, 1, 69–107.

[8] Beckers, S. (1981) A note on estimating parameters of a jump-diffusion process of stock returns", *Journal of Financial and Quantitative Analysis*, **16** (1), 127–140.

[9] Bertoin, J. (1996) *Lévy processes*, Cambridge University Press.

[10] Carr P., Madan D. (1998): Option valuation using the fast Fourier transform, Journal of Computational Finance, 2, 61-73.

[11] Cont, R. & da Fonseca, J. (2002) Dynamics of implied volatility surfaces, *Quantitative Finance*, **2**, 45–60.

[12] Cont R., Bouchaud J.P. & Potters M. (1997): Scaling in financial data: stable laws and beyond, in: B Dubrulle, F Graner & D Sornette (Eds.): *Scale invariance and beyond*, Berlin: Springer.

[13] Eberlein E., Keller U. & Prause, K. (1998) New insights into smile, mispricing and Value at Risk: the hyperbolic model, *Journal of Business*, **71**, No. 3, 371-405.

[14] Eberlein, E. (2001) "Applications of generalized hyperbolic Lévy motion to finance", in: Barndorff-Nielsen, O., Mikosch T. & Resnick S. (Eds.) (2001) *Lévy processes- theory and applications*, Boston: Birkhauser.

[15] El Karoui, N. & Rouge, R. (2000) Pricing via utility maximization and entropy, *Mathematical Finance* 10, no. 2, 259–276.

[16] Engl H., Hanke M., Neubauer A. (1996) *Regularization of inverse problems*, Dordrecht: Kluwer.

[17] Föllmer H. & Schied, A. (2002) *Stochastic finance*, Berlin: De Gruyter.

[18] Frittelli, M. (2000) The minimal entropy martingale measure and the valuation problem in incomplete markets. *Mathematical Finance*, 10, no. 1, 39–52.

[19] Goll T. & Rüschendorf L. (2001): Minimax and minimal distance martingale measures and their relationship to portfolio optimization, *Finance and Stochastics*, **5**, no. 4, 557–581.

[20] Jacod, J. & Shiryaev, A.N. (1987) *Limit theorems for stochastic processes* , Berlin: Springer.

[21] Kallsen, J. (2001): Utility-Based Derivative Pricing. in: *Mathematical Finance - Bachelier Congress 2000*, Berlin: Springer.

[22] Koponen, I. (1995) Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process. *Physics Review E 52*, 1197-1199.

[23] S. G. Kou (2002) A jump diffusion model for option pricing, *Management Science*. Vol. 48, 1086-1101.

[24] Madan D. B., Carr P., & Chang, E.C. (1998). The variance gamma process and option pricing. *European Finance Review*, **2**, 79–105.

[25] Madan, D. (2001) "Financial modeling with discontinuous price processes", in: Barndorff-Nielsen, O., Mikosch T. & Resnick S. (Eds.) (2001) *Lévy processes- theory and applications*, Boston: Birkhauser.

[26] Mancini, C. (2001) Disentangling the jumps from the diffusion in a geometric jumping Brownian motion, *Giornale dell'Istituto Italiano degli Attuari*, Vol LXIV, 19–47.

[27] Merton, R.C. (1976) Option Pricing When Underlying Stock Returns are Discontinuous, *Journal of Financial Economics*, 3 (January-February).

[28] Miyahara, Y. (1999) Minimal Entropy Martingale Measures of Jump Type Price Processes in Incomplete Assets Markets, *Asia-Pacific Financial Markets*, Vol. 6, No. 2, pp. 97-113.

[29] Samperi, D. (2002): Calibrating a Diffusion Pricing Model with Uncertain Volatility: Regularization and Stability, *Mathematical Finance*, Vol. 12, No. 1, pp. 71-87.

[30] Sato, K. (1999) *Lévy processes and infinitely divisible distributions*, Cambridge University Press.

[31] Schoutens, W. (2002) Meixner Processes: Theory and Applications in Finance. EURANDOM Report 2002-004.

[32] Tompkins R. (2001) Stock index futures markets: volatility models and smiles, *Journal of Futures Markets*, **21**, No. 1, 43–78.

A Option pricing by Fourier transform

We recall here the expression, due to Carr & Madan [10] of option prices in terms of the characteristic function of the Lévy process. Due to the special structure of the characteristic function in these models, it is convenient to express option prices in terms of the characteristic function. In particular, for a European call option with log strike k

$$C_T(k) = e^{-rT} E^Q[(e^{s_T} - e^k)^+] \tag{54}$$

where s_T is the terminal log price with density $q_T(s)$. The characteristic function of this density is defined by

$$\phi_T(u) \equiv \int_{-\infty}^{\infty} e^{ius} q_T(s) ds. \tag{55}$$

On the other hand, as remarked above, the characteristic function of the log price is given by the Lévy-Khinchin formula (here we limit ourselves to the compound Poisson case):

$$\phi_T(u) = \exp\{T(-\frac{1}{2}\sigma^2 u^2 + i\gamma(\nu)u + \int_{-\infty}^{\infty}(e^{iux}-1)\nu(x)dx)\} \tag{56}$$

$$\gamma(\nu) = r - \frac{\sigma^2}{2} - \int_{-\infty}^{\infty}(e^x - 1)\nu(x)dx \tag{57}$$

In some important cases this characteristic function is known analytically; otherwise one can discretize the Lévy measure and use (in the compound Poisson case) the Fast Fourier transform to compute the characteristic function.

Following Carr and Madan [10] we use Fourier transform methods to evaluate the expression (54) for a given Levy measure. To do so we observe that although the call price as a function of log strike is not square integrable, the time value of the option, that is, the function

$$z_T(k) = E[(e^{s_T} - e^k)^+] - (1 - e^{k-rT})^+ \tag{58}$$

equal to the price of the option (call or put) which is for given k out of the money (forward), may be square integrable. Here we have assumed without loss of generality that $s_0 = 0$. Let $\zeta_T(v)$ denote the Fourier transform of the time value:

$$\zeta_T(v) = \int_{-\infty}^{+\infty} e^{ivk} z_T(k)dk \tag{59}$$

It can be expressed in terms of the characteristic function of the log-price in the following way. First, we note that since the discounted price process is a martingale, we can write

$$z_T(k) = e^{-rT}\int_{-\infty}^{\infty} q_T(s)ds(e^s - e^k)(1_{k\leq s} - 1_{k\leq rT})$$

Next, we compute $\zeta(v)$ by interchanging integrals

$$\zeta_T(v) = e^{-rT}\int_{-\infty}^{\infty} dk \int_{-\infty}^{\infty} ds e^{ivk} q_T(s)(e^s - e^k)(1_{k\leq s} - 1_{k\leq rT})$$

$$= e^{-rT}\int_{-\infty}^{\infty} q_t(s)ds \int_s^{rT} e^{ivk}(e^k - e^s)dk$$

A sufficient condition allowing us to justify the interchange of integrals is that the stock price have a moment of order $1+\alpha$ for some positive alpha or

$$\exists\, \alpha > 0 : \quad \int_{-\infty}^{\infty} q_T(s)e^{(1+\alpha)s}ds < \infty \tag{60}$$

We can write for the inner integral:

$$\int_s^{rT} |e^k - e^s|dk \leq e^{rT} - e^s, \quad \text{if} \quad rT \geq s \tag{61}$$

and

$$\int_{rT}^{s} |e^k - e^s| dk \leq e^s(s - rT)1_{s>rT}, \quad \text{if} \quad rT < s$$

We see that under the condition (60) both expressions when multiplied by $q_T(s)$ are integrable with respect to s and we can apply Fubini's theorem to justify the interchange. The inner integral is computed in a straightforward fashion, and after computing the outer integral for some terms and reexpressing it in terms of the characteristic function of the log stock price, we obtain

$$\zeta_T(v) = \frac{e^{-rT}\phi_T(v-i) - e^{ivrT}}{iv(1+iv)} \tag{62}$$

The martingale condition guarantees that the numerator is equal to zero for $v = 0$. Under the condition (60), we see that the numerator becomes an analytic function and the fraction has a finite limit for $v \to 0$. The option prices can now be found by inverting the Fourier transform:

$$z_T(k) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-ivk}\zeta_T(v)dv \tag{63}$$

B Existence of a solution for the discretized problem

In this section we present a proof of proposition 3. First, we will establish the continuity of J on $\mathcal{L}_\pi$. Then we will establish a lower bound for $H(\nu)$ which will enable us to show that the minimum of $J(\nu)$ is reached on a bounded subset of $\mathcal{L}_\pi$. Finally, we will show that for α large enough this minimum is unique.

Continuity of the relative entropy functional in the discretized case. Consider the functional (23) for a Lévy measure which belongs to the class (34). In this case, all integrals become finite sums and the relative entropy takes the following form:

$$H(\nu) = \frac{T}{2\sigma^2}\left\{\sum_{i=1}^{N}(e^{x_i}-1)(\nu_i - \nu^P(x_i))\right\}^2 + T\sum_{i=1}^{N}\left\{\nu_i \ln\frac{\nu_i}{\nu^P(x_i)} + \nu^P(x_i) - \nu_i\right\} \tag{64}$$

where we denote $\nu_i \equiv \nu(x_i)$ The first term is continuous because it is a continuous function (square) of a finite sum of continuous (linear) functions of ν_i. To treat the second term consider the one-dimensional function $f(x) = x\ln(x/x_0) + x_0 - x$. It is continuous for $x \geq 0$ if we take by definition $f(0) = x_0$ (in fact, it is even uniformly continuous on this set). Thus, the second term is continuous for $\nu_i \geq 0, \ \forall i$.

Continuity of the pricing error. To establish the continuity of the pricing error functional, it is sufficient to prove that one option price is a continuous functional of the Lévy measure or, equivalently, to prove that the time value $z_T(k,\nu)$ defined by equation (58) is continuous for all k. Let ν^1 and ν^2 be two Lévy measures in $\mathcal{L}_\pi$ and take $\varepsilon > 0$. We will prove that there exists $\delta > 0$ such that if $\max_i |\nu_i^1 - \nu_i^2| < \delta$ then $|z_T(k,\nu^1) - z_T(k,\nu^2)| < \varepsilon$. From equations (63) and (62),

$$
\begin{aligned}
|z_T(k,\nu^1) - z_T(k,\nu^2)| &\leq \int_{-\infty}^{\infty} |\zeta_T(v,\nu^1) - \zeta_T(v,\nu^2)| dv \\
&\leq e^{-rT} \int_{-\infty}^{\infty} \frac{|\phi^{\nu^1}(v-i) - \phi^{\nu^1}(v-i)|}{|v|\sqrt{1+v^2}} dv
\end{aligned}
$$

for all k. Equations (56) and (57) entail that $|\phi^\nu(v-i)| \leq e^{rT}$ for all ν and for all $v \in \mathbb{R}$. Hence, one can find A such that

$$
e^{-rT} \int_{|v|\geq A} \frac{|\phi^{\nu^1}(v-i) - \phi^{\nu^1}(v-i)|}{|v|\sqrt{1+v^2}} dv \leq \frac{\varepsilon}{2}
$$

Recalling that $\phi^\nu(u) = \exp T\psi^\nu(v)$ and using once again the upper bound for the modulus of ϕ along with the mean value theorem, we find

$$
\begin{aligned}
e^{-rT} \int_{|v|\leq A} \frac{|\phi^{\nu^1}(v-i) - \phi^{\nu^1}(v-i)|}{|v|\sqrt{1+v^2}} dv &\leq \int_{|v|\leq A} \frac{|\psi^{\nu^1}(v-i) - \psi^{\nu^1}(v-i)|}{|v|\sqrt{1+v^2}} dv \\
\leq \sum_{j=1}^{N} |e^{x_j} - 1||\nu_j^1 - \nu_j^2| \int_{v\leq A} \frac{dv}{\sqrt{1+v^2}} &+ \sum_{j=1}^{N} e^{x_j} |\nu_j^1 - \nu_j^2| \int_{v\leq A} \frac{|e^{ivx_k} - 1|}{|v|\sqrt{1+v^2}}
\end{aligned}
$$

Since all the integrals in the last term are integrals of bounded functions over bounded domains, one can choose $\delta > 0$ such that if $\max_i |\nu_i^1 - \nu_i^2| < \delta$ then the last term is smaller than $\frac{\varepsilon}{2}$ and hence $|z_T(k,\nu^1) - z_T(k,\nu^2)|$ is smaller than ε. This proves the continuity of the pricing functional, uniformly on k.

Lower bound for regularized functional Using the following trivial inequality

$$
x \ln \frac{x}{x_0} + x_0 - x \geq x - x_0(e-1)
$$

we obtain the following bound for the functional $\mathcal{J}(\nu)$:

$$
\mathcal{J}(\nu) \leq \alpha H(\nu) \leq \alpha T \sum_{i=1}^{N} \nu_i - \alpha T(e-1) \sum_{i=1}^{N} \nu_i^P \tag{65}
$$

The regularized functional is thus bounded below by the L^1 norm of ν minus some constant.

Existence of solution for the regularized problem To prove that the regularized problem has a solution, consider a compact $C \subset \mathcal{L}_\pi$ defined by

$$C = \{\nu : \alpha T \sum_{i=1}^{N} \nu_i \leq \alpha T(e-1) \sum_{i=1}^{N} \nu_i^P + \mathcal{J}(\nu^P) + 1\} \tag{66}$$

Since $\mathcal{J}$ is continuous,

$$\exists \nu^* \in L_\pi, J(\nu^*) = \inf_{\nu \in C} \mathcal{J}(\nu)$$

However, using the bound (65) for all $\nu \in L_\pi \setminus C$ we find $\mathcal{J}(\nu) \leq \mathcal{J}(\nu^P) + 1 < \mathcal{J}(\nu^*)$. Hence, ν^* is the solution of the regularized problem.

Uniqueness of the solution for large α Making the additional hypothesis that $\sigma > 0$, we will now show that for any compact K there exists an α_0 such that for any $\alpha \geq \alpha_0$, $\mathcal{J}(\nu)$ is convex on K. Since the size of compact C in (66) decreases when α grows, this entails unicity of the solution for $\alpha > \alpha_0$.

Consider the regularized functional $\mathcal{J}(\nu) = \alpha H(\nu) + \mathcal{E}(\nu)$ where $\mathcal{E}(\nu)$ denotes the sum of squared pricing errors. From equation (64),

$$H^{jk} \equiv \frac{\partial^2 H}{\partial \nu_j \partial \nu_k} = \frac{T}{a}(e^{x_j} - 1)(e^{x_k} - 1) + \frac{T}{\nu_j}\delta_{jk}$$

Suppose that we can prove that the second derivative of $\mathcal{E}(\nu)$ can be represented as follows:

$$\mathcal{E}^{jk} \equiv \frac{\partial^2 \mathcal{E}}{\partial \nu_j \partial \nu_k} = P^{jk} + B^{jk} \tag{67}$$

where P is a positive definite matrix and the elements of B are bounded by some constant c. Let $K = \{\nu : \nu_i \leq \nu_0\}$ for some $\nu_0 > 0$. Then for every $\alpha > Nc\nu_0/T$, $\mathcal{J}$ is convex on K. Indeed, for every vector $\Delta\nu$ and for every $\nu \in K$,

$$\begin{aligned}\sum_{jk} \frac{\partial^2 \mathcal{J}(\nu)}{\partial \nu_j \partial \nu_k} \Delta\nu_j \Delta\nu_k = \frac{\alpha T}{a}(\sum_j (e^{x_j} - 1)\Delta\nu_j)^2 + \alpha T \sum_j \frac{\Delta\nu_j^2}{\nu_j} + \\ \sum_{jk} P^{jk} \Delta\nu_j \Delta\nu_k + \sum_{jk} B^{jk} \Delta\nu_j \Delta\nu_k \geq \\ \alpha T \sum_j \frac{\Delta\nu_j^2}{\nu_j} + \sum_{jk} B^{jk} \Delta\nu_j \Delta\nu_k \geq \alpha T \sum_j \frac{\Delta\nu_j^2}{\nu_0} + Nc \sum_j \Delta\nu_j^2 \geq 0\end{aligned}$$

It remains therefore to prove the representation (67). Without loss of generality we can consider that there is only one option, hence $\mathcal{E}(\nu) = (C(\nu) - C_0)^2$. Furthermore,

$$\frac{\partial^2 \mathcal{E}(\nu)}{\partial \nu_j \partial \nu_k} = 2 \frac{\partial C(\nu)}{\partial \nu_j} \frac{\partial C(\nu)}{\partial \nu_k} + 2(C(\nu) - C_0) \frac{\partial^2 \mathcal{C}(\nu)}{\partial \nu_j \partial \nu_k}$$

The first term in this expression is a positive definite matrix and we must prove the boundedness of the right term. Since the price of an option is always bounded by the current stock price, it is sufficient to check the boundedness of $\frac{\partial^2 \mathcal{C}(\nu)}{\partial \nu_j \partial \nu_k}$ or, equivalently, of $\frac{\partial^2 z(\nu)}{\partial \nu_j \partial \nu_k}$ (here we omit the variable k which is now irrelevant). From formula (63),

$$\frac{\partial^2 z(\nu)}{\partial \nu_j \partial \nu_k} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ivk} \frac{\partial^2 \zeta_v(\nu)}{\partial \nu_j \partial \nu_k} dv$$

thus, for $z(\nu)$ to be bounded $\zeta_v(\nu)$ must be integrable as a function of v uniformly on ν. This is equivalent by formula (62) to uniform integrability of

$$\left| \frac{1}{v(1+iv)} \frac{\partial^2 \phi(v-i)}{\partial \nu_j \partial \nu_k} \right| \tag{68}$$

By Lévy-Khintchine formula,

$$\phi_T(v-i) = e^{T\psi(v-i)} \quad \text{where} \tag{69}$$
$$\psi(v-i) = -\frac{1}{2}\sigma^2 v^2 + \frac{1}{2} iv\sigma^2 + ivr + r - iv \sum_j (e^{x_j} - 1)\nu_j + \sum_j e^{x_j}(e^{ivx_j} - 1)\nu_j$$

Hence,

$$\frac{\partial^2 \phi(v-i)}{\partial \nu_j \partial \nu_k} = T^2 \phi(v-i) \frac{\partial \psi(v-i)}{\partial \nu_j} \frac{\partial \psi(v-i)}{\partial \nu_k}$$

Since

$$\frac{\partial \psi(v-i)}{\partial \nu_j} = -iv(e^{x_j} - 1) + e^{x_j}(e^{ivx_j} - 1),$$

both $\frac{\partial \psi(v-i)}{\partial \nu_j}$ and $\frac{1}{v}\frac{\partial \psi(v-i)}{\partial \nu_j}$ are bounded above by polynomials in v, independent on ν. Formula (69) entails that $|\phi_T(v-i)| \le e^{r - \frac{1}{2}\sigma^2 v^2}$. Therefore, the integral of (68) is bounded by a number which does not depend on ν and the proof is completed.

C Properties of regularized solutions

We present here some properties of the solutions of our regularized problem in the discretized case (i.e. the Lévy measure is concentrated on a discrete grid). This is the only case that is interesting from the point of view of numerical implementation. A proof in the case where Lévy measure is concentrated on a bounded interval may be constructed using the general theory exposed for example in ([16], section 10.6). We shall denote by H the relative entropy functional defined in (25): $H(\nu) = \mathcal{E}(\mathbb{Q}(\sigma,\nu), \mathbb{Q}(\sigma,\nu_0))$. Define $\delta > 0$ as the observational error on the data C^*: $||C^* - C|| \le \delta$ where C^* is the vector of observed option prices and C a vector of arbitrage free ('true') prices.

The solution of (32) is in general not unique due to non-convexity of the pricing functional. It depends continuously on the data in the following sense:

Proposition 4. *Let $\alpha > 0$ and let $\{C^k\}$ and $\{\nu^k\}$ be sequences where $C^k \to C^*$ and ν^k is the solution of problem (32) with C^* replaced by C^k. Then there exists a convergent subsequence of $\{\nu^k\}$ and the limit of every convergent subsequence is a solution of (32).*

Remark: if the solution of (32) is unique this is just the definition of continuity.

Proof. To simplify the notation we write $F(\nu)$ for a set of model prices and $||F(\nu) - C^*||^2$ for the sum of squared differences of model prices corresponding to Lévy measure ν and market prices C^*. Let $\{C_k\}$ be a sequence of data sets converging to C^* and $\{\nu_k\}$ be the corresponding sequence of solutions:

$$\nu_k = \arg\inf\{||F(\nu_k) - C_k||^2 + \alpha H(\nu)\}$$

By construction we have:

$$||F(\nu_k) - C_k||^2 + \alpha H(\nu) \leq ||F(\nu) - C_k||^2 + \alpha H(\nu), \quad \forall \nu \in \mathcal{L}(\mathbb{R}) \tag{70}$$

hence the sequences $||\nu_k||$ and $||F(\nu_k)||$ are bounded. Since we work in a finite-dimensional space, we can find a convergent subsequence $\nu_m \to \nu^*$ of $\{\nu_k\}$. Using the continuity of the pricing functional we have $F(\nu_m) \to F(\nu^*)$. This together with (70) and the continuity of the relative entropy functional implies:

$$\forall \nu \in \mathcal{L}(\mathbb{R}), \quad ||F(\nu^*) - C^*||^2 + \alpha H(\nu) = \lim\{||F(\nu_m) - C_m||^2 + \alpha H(\nu)\} \leq \lim\{||F(\nu) - C_m||^2 + \alpha H(\nu)\} = ||F(\nu) - C^*||^2 + \alpha H(\nu).$$

Hence, we have proven that ν^* is a minimizer of $||F(\nu) - C^*||^2 + \alpha H(\nu)$. □

Let M be the set of Lévy measures ν corresponding to least square solutions which minimize the criterion (30). Assume that

$$\exists \nu \in M, \mathcal{E}(\mathbb{Q}(\sigma, \nu), \mathbb{Q}_0) < \infty \tag{71}$$

Then a minimum-entropy least squares solution is defined as a solution of

$$\inf_{\nu \in M} \mathcal{E}(\mathbb{Q}(\sigma, \nu), \mathbb{Q}_0) \tag{72}$$

The next proposition describes how the solutions of (32) converge towards minimum-entropy least squares solutions as the error level δ decreases.

Proposition 5. *Let*

$$||C^* - C|| \leq \delta$$

and let $\alpha(\delta)$ be such that $\alpha(\delta) \to 0$ and $\delta^2/\alpha(\delta) \to 0$ as $\delta \to 0$. Then every sequence $\{\nu^{\delta_k}_{\alpha(\delta_k)}\}$ where $\delta_k \to 0$ and $\nu^{\delta_k}_{\alpha(\delta_k)}$ is a solution of problem (32) has a convergent subsequence. The limit of every convergent subsequence is a a minimum entropy least squares solution. If the minimum entropy least squares solution is unique, then

$$\lim_{\delta \to 0} \nu^{\delta}_{\alpha(\delta)} = x^*$$

where x^* is the solution of (72).

Proof. Let the sequences $\{\nu^{\delta_k}_{\alpha(\delta_k)}\}$ and δ_k be as above and ν^* be a minimum entropy least squares solution. Then by definition of $\nu^{\delta_k}_{\alpha(\delta_k)}$ and using the triangle inequality we have

$$||F(\nu^{\delta_k}_{\alpha(\delta_k)}) - C^{\delta_k}||^2 + \alpha(\delta_k)H(\nu^{\delta_k}_{\alpha(\delta_k)}) \leq \alpha(\delta_k)H(\nu^*) + ||F(\nu^*) - C||^2 + \delta_k^2$$

Hence, when we pass to the limit

$$\lim_{k\to\infty} ||F(\nu^{\delta_k}_{\alpha(\delta_k)}) - C^{\delta_k}|| = ||F(\nu^*) - C|| \tag{73}$$

Again from triangle inequality and the definition of ν^* we obtain

$$\alpha(\delta_k)H(\nu^{\delta_k}_{\alpha(\delta_k)}) \leq \alpha(\delta_k)H(\nu^*) + 2\delta_k^2$$

This means (dividing by $\alpha(\delta_k)$) that

$$\limsup H(\nu^{\delta_k}_{\alpha(\delta_k)}) \leq \mathcal{H}(\nu^*) \tag{74}$$

and that the sequence $\nu^{\delta_k}_{\alpha(\delta_k)}$ is bounded. Hence we it has a subsequence $\nu^{\delta_m}_{\alpha(\delta_m)}$ converging towards some measure ν as $m \to \infty$. (73) shows that ν is a least squares solutions and from (74) we see that is is a minimum entropy least squares solution. The last assertion follows from the fact that in this case every subsequence of $\nu^{\delta_k}_{\alpha(\delta_k)}$ has a subsequence converging towards ν^*. □