Joint Calibration of Option Pricing Models via Particle Methods

Billy Amzal  
Université Paris-Dauphine, CEREMADE & ENGREF, GRESE & INERIS  
19, Avenue du Maine, 75015 Paris, France  
billy@amzal@polytechnique.org

Yonathan Ebguy  
EHESS & HSBC-DMRG, 103 Av. des Champs Elysées  
75008, Paris Cedex, France  
yonathan.ebguy@ccf.com

Sébastien Roland  
Université d’Evry, Département de Mathématiques & ESSEC, Finance Department  
Bd. François Mitterrand, 91025, Evry Cedex, France  
sebastien.roland@maths-evry.fr

Received (1st June 2005)  
Revised (Day Month Year)

This paper develops a new joint calibration procedure where both implied and historical distributions are simultaneously used. We closely link it to the minimization of measures of risk under uncertainty. We focus on the class of affine jump-diffusion models to derive a parametric formulation of the problem. From both theoretical and practical points of view, we are concerned with the quantitative assessment of the financial risk to build a robust and efficient pricing system. As a loss function, we choose the sum of a classical least-square cost and of a regularizing term of relative entropy. Then, in order to develop a computationally efficient methodology, we translate it into its probabilistic counterpart in a general context of maximization of expected utility (MEU). This leads us to the development of simulation-based algorithms in line with Monte-Carlo Markov Chains (MCMC) methods. To avoid their traditional shortcomings like local mode trapping, we consider an original alternative derived from Interacting Particle Systems (IPS). A new theoretical framework for this method is provided and convergence results are established. This algorithm is applied to simulated data and to a EuroStoxx 50 data set. We extensively discuss these results and interpret them notably in terms of risk aversion and models perception.

Keywords: Calibration; Option Pricing; Jump-Diffusion Model; Particle Systems.

1. Introduction

Stocks or interest-rate derivatives are priced under the risk neutral measure while the statistical measure is used to model the underlying on which the option is written. Classical calibration approaches such as those of Avellaneda [3] and Bates [4]
only resort to the implied measure to fit the observed smile of volatility ignoring information available in the historical measure. Thus, one may wish to build a joint calibration procedure where both implied and historical distributions are simultaneously used. It is worth noting that several attempts have been made to answer these questions as in Chernov-Ghysels [9] or in Eraker [19] but these papers didn’t display satisfying results: the risk premiums were erratic, and their interpretation not obvious. In this article, we make a new attempt to build a bridge between these two worlds. We propose a calibration method using the joint information from market prices of some options as well as of the corresponding stocks on which the positions were written.

We consider an original prediction/calibration problem. For example, at time \( t \), we want to predict some option prices on the coming days. It consists of “predicting” the underlying value with its diffusion under the historical probability and then calculating the corresponding option prices with the risk-neutral actualized expectation. In that sense, the \textit{a posteriori} problem to solve reads as: Given a set of data and a chosen day in this time series, what are the historical and risk-neutral measures that would have implied the best prediction for the option prices quoted on the days following the chosen one. This calibration is a kind of prediction in sample. Formally, this is expressed in the following inverse problem:

**Definition 1.1. (Joint calibration problem)** Given a set of stock prices and liquid call options written on this stock, say \( \{ \bar{y}_t, C(\bar{y}_t, T_i, K_i) \} \) for \( i = 1..I \) and \( t = 1..T \), find two random measures \( Q \in Q \), the risk-neutral one, and \( P \in P \), the historical one, such that the observed option prices are given by the \( P \)-expectation of their theoretical prices on stock values simulated under \( P \):

\[
C(\bar{y}_t, T_i, K_i) = B(t, T_i)E_{t-1}^P[C^Q(Y^P_t, T_i, K_i)|\bar{y}_{t-1}]
\]  

- \( B(t, T_i) \) is the discount factor at time \( t \) with maturity \( T_i \),
- \( C^Q(Y^P_t, T_i, K_i) \) is the theoretical price under the risk-neutral measure \( Q \) of the option considered with strike \( K_i \), maturity \( T_i \) and spot value \( Y^P_t \),
- \( Y^P_t \) is simulated under \( P \) from \( \bar{y}_{t-1} \) at time \( t - 1 \).

The probability measure spaces \( P \) and \( Q \) need to be specified so that both measures are equivalent and that arbitrage opportunities are avoided. This goal is normally met by assigning market price of risk process to the diffusion of the state variable(s) as described by Harrison-Pliska [25] and Harrison-Kreps [24]. This results in the fact that if a process is within the class of affine jump-diffusion under the objective probability measure, the market price of risk specification ensures that it is within the same class under the equivalent martingale measure and vice-versa. The bridge alluded to earlier will also be achieved through the assignment of market prices of risk. This enables one to preserve equivalence of measures and precludes arbitrage opportunities. Therefore, our joint calibration approach will succeed if we manage to separate the impact of the historical parameters and that of the risk-neutral ones. It could help one both to calibrate models and to determine parameters
usually estimated roughly from historical data with a better understanding of their impact and of their evolution.

But, at this stage, the inverse problem (1.1) is ill-posed. Indeed, it is under-determined: the knowledge of a finite numbers of option prices is not enough to characterize the risk neutral measure. There might be no solution or an infinite number of solutions, not necessarily in the class of models we were expecting it to be. Traditionally, in the calibration issues, this obstacle is bypassed by minimizing the in-sample quadratic pricing error. Inspiring ourselves from that idea to obtain a practical solution, we resort to minimizing a measure of risk, or more precisely to minimize the expectation of a loss-function \( L \), defined on \( Q \), the probability measure space to be chosen. The concept of a measure of risk advanced in [2] and refined in [20] and in a large subsequent literature has gained widespread acceptance in the option pricing industry. This approach focuses on the measure of the quantitative risk involved in a financial position. The goal is now to find:

\[
(P, Q) = \arg \min_{(P, Q)} \mathbb{E}^P [L(Q)]
\] (1.2)

This formulation has a mathematical sense if the loss function has good properties. But it could seem numerically difficult to solve. Usual ways of optimizing a function like BFGS descent gradient method would be here non-efficient because of the too numerous calculations of the expectation needed. We thus have to resort to other numerical methods. Following the fast development of computers, simulation techniques have appeared to be a more and more interesting alternative for analytic or algebraic approach of optimization problems. MCMC algorithms have been widely developed and applied this last decade for Bayesian calibration (see [32] for an overview of Bayesian techniques), in finance as well as in many other fields of application. The optimization issue is transformed into a simulation one, the particle algorithms appear to be adapted to our problem. A recently born algorithm inspired by Interacting Particle Systems which doesn’t demand any calculation of the expectation is presented. New powerful convergence properties which help us for the practical implementation are proved.

The remainder of the paper is organized as follows. Section 2 sets the jump-diffusion models considered and defines \( P \) and \( Q \), which gives a parametric formulation of the optimization problem. In Section 2, we also explain the choice of our loss-function, adding to the traditional least-squares cost, an entropy term, regularizing the problem. Section 3 is dedicated to the presentation of a simulation-based optimization algorithm using MCMC approaches and particle methods coupled with simulated annealing. In Section 4, which the reader can skip in first read, we introduce an original formalism for this algorithm to prove new convergence results. In Section 5, this algorithm applied to some specific and largely used option pricing models derived from Section 2, and the numerical results obtained with simulated data are discussed. In section 6, a real-world application is given and risk premiums analysis is derived. Last section concludes.
2. Affine arbitrage pricing models driven by jump-diffusions

We begin with some definitions and properties of affine jump-diffusion models. Moreover, we precise the probability measure spaces $P$ and $Q$ on which the optimization is made and we discuss market price of risk specifications and absence of arbitrage for this class. Finally, we parametrize the joint calibration problem for a specific family of stochastic volatility with jumps models. Throughout this section, $P$ corresponds to the historical probability, and $Q$ to the risk-neutral one.

2.1. Affine jump-diffusions

Let $(\Omega, \mathcal{F}, P)$ be a probability space on which we define the two following independent random processes: a $d$-dimensional standard Brownian motion $(W_t)_{t \geq 0}$ and a $d$-dimensional compound Poisson process $(N_t)_{t \geq 0}$. We also suppose that there is a Markov process $Y$ taking values in some open subset $D$ of $\mathbb{R}^d$ and satisfying the following theorem (see [17]). Let an affine jump-diffusion (AJD) be given by:

$$dY_t = \mu(Y_t, t) dt + \Sigma(Y_t, t) dW_t + d\tilde{N}_t$$

where $\mu : D \rightarrow \mathbb{R}^d$, $\Sigma : D \rightarrow \mathbb{R}^{d \times d}$ and $\tilde{N}_t$ is a compensated Poisson process with compensator $\nu(dy, dt)$. The Lévy measure $\nu$ dictates how jumps occur. In a finite activity setup, jumps arrive with intensity $\lambda = \int_D \nu(dy, dt) : D \rightarrow \mathbb{R}$ with $\lambda < \infty$ and are distributed according to a fixed probability distribution $m$ on $\mathbb{R}^d$, with $\nu(dy, dt) = \lambda m(dy, t)$. In this case we have the following

Theorem 2.1. An AJD process satisfies:

- (affinity) Drift, squared volatility and intensity are all affine such that the determining triplet of characteristics $\kappa = (\mu, \Sigma, \nu)$ writes as follows:

$$\mu = k_0(t) + k_1(t) \cdot Y$$

$$\Sigma \Sigma^T = h_0(t) + h_1(t) \cdot Y$$

$$\nu = l_0(t) + l_1(t) \cdot Y$$

- (ode-s) Coefficients are such that solutions $\beta$ and $\alpha$ to the following system of ordinary differential equations exist:

$$\beta'(t) = p_1(t) - k_1(t)^T \beta(t) - \frac{1}{2} \beta(t)^T h_1(t) \beta(t) - l_1(t) \int_D e^{\beta(t) \cdot y} dm(y) - 1$$

$$\alpha'(t) = p_0(t) - k_0(t)^T \beta(t) - \frac{1}{2} \beta(t)^T h_0(t) \beta(t) - l_0(t) \int_D e^{\beta(t) \cdot y} dm(y) - 1$$

with boundary conditions $\beta(T) = u$ and $\alpha(T) = 0$.

Then, for each $u \in \mathbb{R}^d$, the discounted characteristic function process

$$\phi^u(u, Y_t, t, T) = \mathbb{E}_t^Q \left( e^{-r(T-t)} e^{u \cdot Y_T} | \mathcal{F}_t \right)$$

(2.3)
has exponential affine form in $X$, namely

$$\phi^X(u, Y_t, t, T) = \exp (\alpha(t) + \beta(t) \cdot Y_t)$$

Let us consider a contingent claim on $Y$ whose profile is given by a final cash amount $g(Y_T) \equiv g(Y_T, \cdot)$ and whose value at time $t$ is:

$$P_t = \mathbb{E}_t^Q \left[ e^{-r(T-t)} g(Y_T, T, K) \right]$$

The general result stated above allows a direct pricing of such claim, in the case of affine pay-off functions in the factor $Y$, via Fourier transform methods. In fact, we can use the general result in Theorem 2.1 to compute characteristic function distributions and then numerically invert them to recover the corresponding density function. This leads to a general call option formula for AJD stock price models:

**Proposition 2.1.** Let us consider a call option on a function $g$ of the underlying factor expiring at time $T$ and striking at price $K$. For any non-decreasing function $h$, the price of the call is then:

$$P_t = \mathbb{E}_t^Q \left[ e^{-r(T-t)} (g(Y_T) - K)^+ \right] = f_g(-h(K)) - K f_1(-h(K))$$

(2.4)

with $I$ the identity function and where:

$$f_g(y) = \mathbb{E}_t^Q \left[ e^{-r(T-t)} g(Y_T) I_{\{ -h(g(Y_T)) \leq y \}} \right]$$

can be calculated by inverting its Fourier transform as

$$F[f_g](u) = \int \mathbb{E}_t^Q \left[ e^{-r(T-t)} g(Y_T) I_{\{ -h(g(Y_T)) \leq y \}} \right] dy$$

$$= \mathbb{E}_t^Q \left[ e^{-r(T-t)} g(Y_T) \int \mathbb{E}_t^Q \left[ e^{iux} I_{\{ -h(g(Y_T)) \leq y \}} \right] dy \right]$$

$$= \mathbb{E}_t^Q \left[ e^{-r(T-t)} g(Y_T) e^{iu(-h(Y_T))} \right]$$

where $h$ has to be chosen such that the expectation (2.4) is of the form (2.3).

### 2.2. Market price of risk specifications and absence of arbitrage

The following proposition is a consequence of Girsanov transformation for semi-martingales ([27]). In all the sequel, we use the notation $f * \nu$ meaning that we integrate the function $f$ with respect to the measure $\nu$.

**Proposition 2.2.** Let $Y$ be an AJD with $\mathbb{P}$-characteristics $(\mu^P, \Sigma^P, \nu^P)$. For any probability measure $Q \ll \mathbb{P}$, there exist a predictable function $\Lambda_3 > 0$ and a predictable $\mathbb{R}^d$-valued process $\Lambda_1$ such that $Q$-characteristics of $Y$ are given by:

$$\mu^Q = \mu^P + \Sigma^P \cdot \Lambda_1 + (\Lambda_3 - 1) * \nu^P$$

$$\Sigma^Q = \Sigma^P \cdot \Lambda_3$$

$$\nu^Q = \Lambda_3 \cdot \nu^P$$

(2.5)
A_1 and A_3 are called the Girsanov quantities of Q with respect to P relative to Y. Intuitively, A_3 describes how the jump distributions of Y changes when we turn from the historical measure P to the risk-neutral one Q and A_1 together with A_3 determines the change in drift.

To give a deeper insight of market price of risk specifications, one can express the density process Z^Q of Q with respect to P explicitly in terms of A_1 and A_3. If we denote the Doleans-Dade exponential E, we have the following proposition based on weak representation property for semi-martingales (see [27]):

**Proposition 2.3.** Let Y be an affine jump-diffusion. If Q ≪ P with Girsanov quantities A_1 and A_3, the density process of Q with respect to P is given by

\[ Z^Q_t = E^P \left( N^Q_t \right) \]  

with:

\[ N^Q_t = (A_1 \cdot Y^c + (A_3 - 1) \ast (m^P - \nu^P))_t \]  \hspace{1cm} (2.6)

If P exists and is solution of a martingale problem for the diffusion (2.1), existence of the quantities A_1 and A_3 is sufficient neither for the existence of the implied probability measure Q nor for its equivalence to P. However, these two properties are jointly entailed by the following necessary and sufficient condition:

\[ E^P_t \left[ E \left( N^Q \right) \right] = 1 \]  \hspace{1cm} (2.7)

The existence of an equivalent martingale measure implies the absence of arbitrage opportunity ([14]). Our specification for market price of risk in affine jump-diffusion models precludes arbitrage opportunities and the models under both measures are of the same class.

**Theorem 2.2.** Let \((\Omega, \mathcal{F}, P)\) be a probability space containing a d-Brownian motion \((W^P_t)_{t \geq 0}\) and a d-compensated Poisson process \((\tilde{N}^P_t)_{t \geq 0}\) such that there exists a stochastic process \((Y_t)_{t \geq 0}\) satisfying (2.1). Then, there exists a measure Q equivalent to P such that:

\[ dY_t = \mu^Q(Y_t, t) dt + \Sigma^Q(Y_t, t) dW^Q_t + d\tilde{N}^Q_t \]  \hspace{1cm} (2.8)

such that \((W^Q_t)_{t \geq 0}\) is a Q-Brownian motion and \((\tilde{N}^Q_t)_{t \geq 0}\) a Q-compensated Poisson process.

### 2.3. Affine pricing jump-diffusion

In \((\Omega, \mathcal{F}, P)\) we suppose that Y is the price process, strictly positive, of a security that pays no dividend. The state process is Z = (Y, V)^T where V is the volatility process. Under the statistical measure P, the dynamics of Z characterized by the triplet \((\mu^P, \Sigma^P, \nu^P)\) is given by:

\[ dZ_t = Z_t \left[ \mu^P(\gamma) dt + \Sigma^P(\gamma) dW_t + \int N^P(dz, dt) - \nu^P(\gamma)(dz, dt) \right] \]  \hspace{1cm} (2.9)
where $W$ and $N$ are respectively two-dimensional Brownian motion and Poisson process and $
u^\mathbb{P} (dz, dt) = \lambda^\mathbb{P} dm^\mathbb{P} (z, t)$ is the random measure attached with jumps which arrive with intensity $\lambda$ and mark according to the probability distribution $m^\mathbb{P}$. The probability measure $\mathbb{P}$ is completely defined by the parameters $\gamma^\mathbb{P}$ of the Brownian motion and the Lévy measure, so that we have the set of $\mathbb{P}$-parameters:

$$
\gamma^\mathbb{P} = (\gamma (W^\mathbb{P}), \gamma (\nu^\mathbb{P}))
$$

Given a specification of $(\mu^\mathbb{P}, \Sigma^\mathbb{P}, \nu^\mathbb{P})$ such that a solution of (2.9) exists, we may consider the existence of an equivalent martingale measure $\mathbb{Q}$ by specifying market price of risk $\Lambda_1$ and $\Lambda_3$ which satisfies equations (2.5). Under this measure, $Z$ is an affine jump-diffusion with triplet $(\mu^\mathbb{Q}, \Sigma, \nu^\mathbb{Q})$ such that:

$$
dZ_t = Z_{t-} \left[ \mu^\mathbb{Q}_t (\gamma) \, dt + \Sigma_t (\gamma) \, dW^\mathbb{Q}_t + \int N^\mathbb{Q} (dz, dt) - \nu^\mathbb{Q}_t (\gamma) (dz, dt) \right] \tag{2.10}
$$

Similarly, the measure $\mathbb{Q}$ can be specified by a set of Brownian motion and Lévy measure parameters, so that we define the set of $\mathbb{Q}$-parameters:

$$
\gamma^\mathbb{Q} = (\gamma (W^\mathbb{Q}), \gamma (\nu^\mathbb{Q}))
$$

By Girsanov’s theorem for jump-diffusion processes, cf. Proposition 2.2, we define the market price of risk $\Lambda = (\Lambda_1, \Lambda_3)$ as solutions of the following system:

$$
\mu^\mathbb{Q}_t = \mu^\mathbb{P}_t (\gamma) + \Sigma^\mathbb{P}_t (\gamma) \cdot \Lambda_1 + \int (\Lambda_3 (z) - 1) \, dm^\mathbb{P} (z, t) \tag{2.11}
$$

$$
W^\mathbb{Q}_t = W^\mathbb{P}_t - \int \Lambda_1 \, dt \tag{2.12}
$$

$$
\lambda^\mathbb{Q} dm^\mathbb{Q} (z,t) = \Lambda_3 (z) \cdot \lambda^\mathbb{P} dm^\mathbb{P} (z, t) \tag{2.13}
$$

This presentation of a specific affine jump-diffusion framework raises attention to a number of common pricing models. The general formulation (2.9) or (2.10) includes the pioneering models of Black-Scholes ([5]) or Merton ([29]) in the univariate case. Another widely studied specification is the class of stochastic volatility models with the derivation of Heston ([26]), Bates ([4]) and in a jump-diffusion approach those of Eraker ([19]).

Finally, to cast specifications (2.9) and (2.10) into an estimation framework, we define the set of parameters involved in our optimization procedure:

$$
\gamma = (\gamma^\mathbb{Q}, \Lambda) \epsilon \mathbb{R}^m
$$

At this stage, the objective is to estimate the model-dependent set of parameters $\gamma$. What remains to be done to have a full formulation of the problem is to choose a relevant loss-function.
2.4. Towards a well-posed optimization problem

2.4.1. Regularizing by relative entropy

A natural loss-function \( L(\gamma) \) interesting for the risk minimization (1.2) is the quadratic pricing error. The problem would then become: Given a data set \( D = \{ y_t, \mathcal{C}(y_t, T_i, K_i), t = 1..T, i = 1..I \} \), find:

\[
\gamma^* = \arg \min_{\gamma \in \mathbb{R}^m} L(\gamma)
\]

with:

\[
L(\gamma) = \sum_{i,t} \omega_i \mathbb{E}_{Y_t | \gamma, y_{t-1}} \left| \mathcal{C}^\gamma_t (Y_t, T_i, K_i) - \mathcal{C}(y_t, T_i, K_i) \right|^2
\]

where:

- \( \omega_i \) are the weights to be chosen by the decision maker (see Section 5),
- \( Y_t \) is simulated under \( \mathbb{P}^\gamma \), the objective measure given \( \gamma \) knowing \( y_{t-1} \),
- \( \mathcal{C}^\gamma_t (Y_t, T_i, K_i) \) is the theoretical price at time \( t \) with spot \( Y_t \) of the option of maturity \( T_i \) and strike \( K_i \) under \( \mathbb{Q}^\gamma \), the pricing measure induced by \( \gamma \).

But, as Cont and Tankov notice it in [11] from which this subsection is widely inspired, this formulation, though giving a statistical sense to the problem, doesn’t resolve the uniqueness and stability issues. Indeed, \( L \) is not convex, so many local minima might exist or some flat directions might make the solution get unstable. It is an ill-posed problem. The usual way to cope with that issue is to introduce as in [11], a penalization term, namely the relative entropy of the pricing measure \( \mathbb{Q}^\gamma \) with respect to some prior model \( \mathbb{Q}_0 \) explicitly given by:

\[
I_t (\mathbb{Q}^\gamma | \mathbb{Q}_0) = \mathbb{E}^{\mathbb{Q}^\gamma} \left[ \log \frac{d\mathbb{Q}^\gamma}{d\mathbb{Q}_0} | \mathcal{F}_t \right]
\]

The relative entropy has a few interesting properties which make it relevant to use as a penalization term:

- Financial issue: if \( \mathbb{Q}^\gamma \) is not absolutely continuous with respect to the prior, \( I_t (\mathbb{Q}^\gamma | \mathbb{Q}_0) \) becomes infinite. Thus, if the prior is well chosen, we can impose good properties to \( \mathbb{Q}^\gamma \). For example, we might take an auxiliary simpler diffusion model.
- Numerical aspect: \( I_t (\mathbb{Q}^\gamma | \mathbb{Q}_0) \) is convex in the different parameters, so the penalization term has a “convexification impact” on the surface to optimize which brings stability to the solution(s).
- Information-theoretic foundation: minimizing \( I_t (\mathbb{Q}^\gamma | \mathbb{Q}_0) \) corresponds to adding the least possible information to the prior to fit in the best way with the option prices (and implicitly with the historical evolution of the underlying). So it introduces a tradeoff between the accuracy of the fit (information contained in option prices) and the numerical stability of the results (information contained in the prior). It could therefore also be interesting to take the objective measure as a prior.
The following result shows that in the case where the measures are generated by affine jump-diffusions, the relative entropy can be expressed in terms of the Girsanov parameters and $Q_0$-characteristics of $X = \log(Y)$.

**Proposition 2.4.** If $Q^\gamma \ll Q_0$ with Girsanov quantities $\Lambda_1$ and $\Lambda_3$, the entropy process of $Q^\gamma$ with respect to $Q_0$ is explicitly given by:

$$I_t(Q^\gamma|Q_0) = \frac{1}{2} \mathbb{E}^{Q^\gamma} \left[ \Lambda_1 \cdot \Sigma \right] + \mathbb{E}^{Q^\gamma} \left[ f(\Lambda_3) * \nu^{Q_0} \right]$$

where $f(y) = y \log (y) - (y - 1)$.

**Proof.** Denote by $T = T^{Q^\gamma} = \mathcal{E}(N)$ the density process with respect to $Q_0$. The canonical decomposition of the $Q_0$-submartingale $T \log(T)$ is $T \log(T) = M + A$ with:

$$M = \int Z_-(1 + \log T_-) dN + (T_- f(\Lambda_3)) * (m^{Q_0} - \nu^{Q_0})$$

$$A = \frac{1}{2} \int T_- d \langle N^c \rangle + (T_- f(\Lambda_3)) * \nu^{Q_0}$$

where $M$ is a local $Q_0$-martingale and $A$ is predictable and of finite variation. The quadratic variation $\langle N^c \rangle_t = \Lambda_1 \cdot \Sigma$ is the same under both measures $Q_0$ and $Q^\gamma$. Hence:

$$I_t(Q^\gamma|Q_0) = \mathbb{E}^{Q_0} \left[ T_t \log T_t \right] = \mathbb{E}^{Q_0} \left[ A_t \right]$$

$$= \frac{1}{2} \mathbb{E}^{Q^\gamma} \left[ \Lambda_1 \cdot \Sigma \right] + \mathbb{E}^{Q^\gamma} \left[ f(\Lambda_3) * \nu^{Q_0} \right]$$

2.4.2. **The well-posed optimization problem**

Let us consider the following regularized problem, where $\alpha$ is the weight given to the accuracy (or to the stability):

$$\gamma^* = \arg \min_{\gamma \in \mathbb{R}^m} L'(\gamma)$$

with:

$$L'(\gamma) = \sum_{i,t} \omega_i \mathbb{E}^{Y_t(\gamma, \gamma_{t-1})} \left[ C^\gamma_t(Y_t, T_i, K_i) - C^t(Y_t, T_i, K_i) \right]^2 + \alpha I_t(Q^\gamma|Q_0)$$

The role of $\alpha$ is clearly important and its value should depend on the data used (which govern the shape of the function to be optimized) and on the loss of precision due to the introduction of the entropy term that we allow. This corresponds to what is called an *a posteriori* choice of $\alpha$. We won’t detail here the way to determine a good value for $\alpha$, the interested reader will refer to [11]. Their determination is based on the Morozov discrepancy principle described in [18]. This way to determine $\alpha$ gives the convergence of the solution towards a minimum entropy least squares solution when the error level allowed alluded to earlier tends to zero.

There are two main advantages for this new formulation:
• The first one is that it transforms the ill-posed problem into a well-posed one. The existence of the solution is easy to prove. Let us give here an idea of the proof. We assume simple conditions on the jumps and on the normal laws of the Brownians to prevent the underlying level from being multiplied by more than 100 from one day to another (which is always satisfied in practice). Then the corresponding prices of the options are bounded. So the least squares term is bounded. Besides, the entropy term explodes as soon as the parameters are far from the prior ones. So the infimum of $L'$ is in a compact set. As $L'$ is continuous with respect to $\gamma$, this infimum is a minimum. For $\alpha$ big enough, we could show that this solution is unique.

We now seek $\gamma$ in a compact set $\Gamma \subset \mathbb{R}^m$

• In the neighborhood of the minimum, the surface to explore is more convex which helps one to locate the minimum sought.

Now, we can say that the problem is well-posed in the two senses of the term: theoretically because it has an “admissible” solution and practically because it will be easier to solve.

3. A simulation-based calibration algorithm

In this section, we study the simulation-based approach for the optimization problem previously exposed. To cast the problem from its numerical into its probabilistic formulation, we turn it into the general statistical context of maximization of expected utility (MEU). For some reasons rising from the methodology exposed below, the optimization of interest is interpreted in terms of the maximization of a utility function $U$ instead of the minimization of the loss function $L$.

3.1. The general context of maximization of expected utility

In this subsection, we recall the general framework in which we embed our optimization problem. Our idea is based on a practical point of view of calibration: we want to interpret such calibration or inference problems as a decision one. Indeed, given our data set $\mathcal{D} = \{\tilde{y}_t, \bar{C}(\tilde{y}_t, T, K), t = 1..T, i = 1..I\}$ a decision maker defines a utility function to be maximized, namely $U_\mathcal{D}(\gamma)$ or more simply $U(\gamma)$, such that:

- if we denote $u$ the joint utility with:
  
  $u(\gamma, y) = M - \left( \sum_{i,t} \omega_i \left| C_\gamma^T (y_t, T, K_i) - \bar{C}(\tilde{y}_t, T, K_i) \right|^2 + \alpha I_t (Q | Q_0) \right)$

- if the parameter $\gamma$ is the chosen one for calibration in $\Gamma$

- if $y \in \mathcal{Y}$ is the vector of “predictive” data ($y_t$, $t = 1..T$) drawn from density distribution $p_\mathcal{D}(y | \gamma)$, more simply denoted by $p(y | \gamma)$ and defined as:
  
  $p(y | \gamma) = \delta_{\tilde{y}_1} \otimes p(y_2 | \tilde{y}_1, \gamma) \otimes ... \otimes p(y_T | \tilde{y}_{T-1}, \gamma)$
where \( \mathcal{Y} \), the space of underlying variables, is taken without loss of generality as \( \mathbb{R}^+ \) (for stochastic volatility models, the dimension is 2)

- if the constant \( M \) is such that \( u \) is strictly positive,

then:

\[
\gamma^* = \arg \max_{\gamma \in \Gamma} U(\gamma) = \arg \max_{\gamma \in \Gamma} E_{p(y_i|\gamma, y_{i-1})}[u(\gamma, y)]
\]

and from previous section, we get that \( u \) is bounded and continuous over \( \Gamma \).

We are now facing a utility maximization problem, in the statistical sense of it. However, the utility function \( U(\gamma) \) is very costly to compute in our case, and classical optimization methods (like gradient methods) would not be practicable. We propose then a general alternative to solve this problem.

### 3.2. MCMC and particle simulation methods

#### 3.2.1. MCMC for stochastic optimization

The key idea, originally developed by Müller ([30]), is to consider \( U \) as a pdf (up to a multiplicative constant) over the decision space \( \Gamma \) and to generate a sample from this distribution, the mode of which corresponds to the optimal decision \( \gamma^* \).

This is made possible by the fact that \( u \) is bounded, positive and continuous. As \( U \) is costly to compute in practice, we need to introduce an augmented probability distribution \( h \) over \( (\Gamma, \mathcal{Y}) \):

\[
h(\gamma, y) \propto u(\gamma, y)p(y|\gamma)
\]

We can compute \( h \) explicitly, up to a normalizing constant, since \( p(y|\gamma) \) and \( u(\gamma, y) \) are both easy to compute. By definition, the marginal of \( h \) on \( \gamma \) is proportional to \( U \), as:

\[
\int h(\gamma, y)dy \propto U(\gamma)
\]

We are now in the context where MCMC simulation methods apply. So we can generate a sample from \( h \) and then from \( U \). As we are interested in the mode of \( U \), this method is not very efficient especially since dimensionality of our problem is relatively high (\( \text{dim } \Gamma \) can be up to 10) and since our surface is quite complex. As a consequence, we need to improve the mode search.

#### 3.2.2. Simulated annealing

A classical improvement of this approach is given by the simulated annealing algorithm that we can easily adapt. The idea (used in [30] or in [6]) is now to simulate a sample from \( U^J \), where \( J \) is a large integer. This will obviously sharpen the top of utility surface and concentrate simulations closer to the mode.
We could use the same fixed value of \( J \) for all iterations, but it is not efficient in high dimensional cases. We can also use a "cooling" schedule that makes \( J(n) \) increase up to \(+\infty\) when \( n \) tends to \(+\infty\). If the simulation method is of Metropolis-Hastings type, we know from [21] that the asymptotic condition that ensures convergence to the mode is:

\[
J(n) < \frac{\log n}{m(\sup_{\gamma} U(\gamma) - \inf_{\gamma} U(\gamma))}
\]

where \( m \) is the dimension of \( \Gamma \). Following a similar idea in our MEU context, we introduce a new joint augmented distribution \( h_J \) on \( \Gamma \times \mathcal{Y}^J \) defined as:

\[
h_J(\gamma, y_1, \ldots, y_J) \propto \prod_{j=1}^{J} u(\gamma, y_j) p(y_j | \gamma)
\]

with independence on \( Y_j \) variables, we keep the key property:

\[
\int \ldots \int h_J(\gamma, y_1, \ldots, y_J) \, dy_1 \ldots dy_J \propto \mathcal{U}^J(\gamma)
\]

so that a sample from \( h_J \) would marginally give us a sample from \( \mathcal{U}^J \). The limits of MCMC methods, especially in high dimension and when implemented with simulated annealing, are definitely problems of local modes which can trap Markov chains. In our joint calibration procedure, the utility surface is rather complicated and high-dimensional, we are very concerned by such limits. We propose here an original algorithm based on interacting particles systems, exploiting all the improvements exposed above, for a better exploration of \( \Gamma \), and then for a more efficient marginal mode research of \( h_J \).

3.2.3. A particle approach

This section describes an original alternative to standard MCMC for a MEU problem, first proposed in [1], and as far as we know, not exploited in the field of finance. The method is based on the recent developments of particles filters (see [16]) and population Monte Carlo (see [7]) simulations. To simulate a sample from \( h_J \), we no longer simulate one Markov chain \((\gamma^{(n)}, y_j^{(n)})_{j=1}^{J} \) like in Müller's algorithm, but...
we generate instead $N$ “parallel” chains $(\gamma_i^{(n)}, y_{i,j}^{(n)})_{i=1..N,j=1..J}$. Using the vocabulary from sequential MCMC theory, each couple $(\gamma_i^{(n)}, y_{i,j}^{(n)})$ is called a particle, and the set of $N$ Markov chains is called an Interacting Particles System. Our interest is not to produce a sample to approximate the target distribution but rather to simulate particles close to the modes. It can be done in three steps:

- the intuitive idea is to generate, at each iteration $n$, an approximated weighted sample from $h_J$ by “importance sampling” (see [22]).
- a selection procedure is performed to duplicate particles closer to the modes of the target distribution while eliminating the others. A standard selection procedure can be a “sampling importance resampling” scheme (as described in [33] or [34]). This selection procedure has been widely studied and applied in the literature about particles methods for sequential MCMC (see [16]). Note that other ways of selection are possible (see [8] or [28]) but we won’t discuss them here.
- As usual for this type of algorithm, an independent Markov step with $h_J$ as target distribution is also added for each particle, to avoid degeneracy problems.

At each step (importance sampling, resampling, and Markov step) and for any iteration $n$, this would generate a sample $(\xi_i^{(n)})_{i=1..N}$ from $h_J$ such that as $N \to +\infty$, we have the approximation:

$$\frac{1}{N} \sum_{i=1}^{N} \phi(\xi_i^{(n)}) \simeq \int \phi(\xi)h_J(\xi) d\xi$$

for any measurable and bounded function $\phi$. The interest of this approach is to get a rich sample from $h_J$. The obvious drawback is that this iterative scheme for fixed $J$ and $N$ would cumulate noise so that the approximation would worsen with iterations. This point has been underlined in developments of non-sequential population MCMC algorithms (see [7] or [10]).

However the interest of such iterative algorithms is fully recovered when the target distribution changes with iterations, like in sequential MCMC or particle filters algorithms. In our case, this holds since we add a simulated annealing effect, so that $J$ grows with iterations $n$. Note that in this context, we don’t have anymore constrains about the form of the cooling function $J(n)$ to get the convergence of the algorithm, as we will see it in the convergence section. When no confusion is possible, we will keep writing $J$ instead of $J(n)$, for lighter notation.

### 3.3. A general particle algorithm with simulated annealing

This yields to an original algorithm for optimization that encompasses simulated annealing to a particle approach, as first formulated in [1]. As Markov step, a common and convenient choice we will make in the practical aspect of this paper is
a Metropolis-Hastings step, since it is easy to implement in practice. We will afford a simplification in the theoretical part (see Section 4.2). We need to choose two random walk jump functions over Γ: \( q_{1,n} \) as transition kernel for the importance sampling step and \( q_{2,n} \) for the Metropolis-Hastings jump. These functions can be widely different: for example \( q_{1,n} \) could have a smaller variance to allow a small scale exploration at the importance sampling step, and a larger scale one at the Metropolis-Hastings step. We will precise these choices in Sections 4 and 5.

With these notations, the weights for importance resampling (multinomial law) are defined as:

\[
 w_i^{(n)} \propto \frac{\prod_{j=1}^J u(\gamma_i^{(n)}, \gamma_j^{(n)})}{q_{1,n}(\gamma_i^{(n-1)}, \gamma_j^{(n)})}
\]

The detailed algorithm can be written as follow:

**Algorithm 3.1. (General Particle Optimization Algorithm)**

1. **Initialization:** Start with a sample \((\gamma_i^{(0)})_{i=1,N}\) at \( t = 0 \). Set \( J = J(0) \).
2. **Importance Sampling Step:** For each \( i = 1..N \), simulate \( \tilde{\gamma}_i^{(1)} \) from \( K_{1,1}(\gamma_i^{(0)}, \cdot) \) and \( J \) independent data sets \((\gamma_j(i))_{j=1..J}\) from \( p(y|\tilde{\gamma}_i^{(1)}) \) for \( j = 1..J \). Compute \( u_i^{(1)} = \prod_{j=1}^J u(\gamma_i^{(1)}, \gamma_j^{(1)}) \) and the normalized weights \( w_i^{(1)} \propto u_i^{(1)}/K_{1,1}(\gamma_i^{(0)}, \gamma_i^{(1)}) \).
3. **Selection Step:** Resample \((\gamma_i^{(1)}, \ldots, \gamma_N^{(1)})\) from \((\gamma_i^{(1)}, \ldots, \gamma_N^{(1)})\) according to a multinomial distribution with weights \( w_i^{(1)} \), note \( u_i^{(1)} \) the corresponding utilities.
4. **Metropolis-Hastings Step:** For each \( i = 1..N \), simulate \( \gamma_i^{(1)} \) from \( K_{2,1}(\tilde{\gamma}_i^{(1)}, \cdot) \) and \( J \) independent data sets \((\gamma_j(i))_{j=1..J}\) from \( p(y|\gamma_i^{(1)}) \) for \( j = 1..J \). Compute \( u_i^{(1)} = \prod_{j=1}^J u(\gamma_i^{(1)}, \gamma_j^{(1)}) \) and the acceptance rates \( \alpha_i = \min(1, (u_i^{(1)} K_{2,1}(\gamma_i^{(1)}, \gamma_i^{(1)}))/u_i^{(1)}/K_{2,1}(\gamma_i^{(1)}, \gamma_i^{(1)}))) \). Set \( \gamma_i^{(1)} = \gamma_i^{(1)} \) with probability \( \alpha_i \) and \( \gamma_i^{(1)} = \gamma_i^{(1)} \) with probability \( 1 - \alpha_i \).

We loop the last three steps until \( J \) is sufficiently large to allow mode determination.

### 3.4. A particular case: the resampling Markov algorithm

Just like in [1], we can notice an important particular case of this general algorithm. By removing the importance sampling step, we can only consider loops of multinomial resampling and Markov steps. Indeed, if one gets at the beginning of loop \( n \) an approximated sample from \( \pi_n \), it becomes an approximated sample from \( \pi_{n+1} \) after resampling and Markov rejuvenating. We assume that \( J(n) > J(n-1) \) to give a sense to the resampling, which means that, in this case, the cooling schedule is at least linear. In this setup, the resampling weights have a much simpler form:

\[
 w_i^{(n)} \propto \prod_{j=J(n)+1}^{J(n)} u(\gamma_i^{(n)}, \gamma_j^{(n)})
\]
where $y^{(n)}_{ij}$ are additional independent draws from $p(y|\gamma^{(n)}_i)$. The Markov transition (denoted $K_n$ in this case) is the same as in the general algorithm, we take the Metropolis-Hastings Markov step to describe the algorithm below. An obvious advantage of this Resampling-Algorithm is a substantial saving of computation time, which may be used to improve the Markov kernel (like using adaptive Markov transitions). A practical drawback lies in the exploration of the utility surface in the most complex multimodal cases. However, we will see that this algorithm is also theoretically more “stable”, as we will show an uniform convergence theorem for it, provided a linear form of the cooling schedule.

Algorithm 3.2. (Resampling-Markov Algorithm)

1. **Initialization:** Start at $t = 0$ with a sample $(\gamma^{(0)}_i, y^{(0)}_{ij})_{i=1..N}$ drawn by importance sampling like in step 1 on the general algorithm. Set $J = J(0) = 1$.

2. **Reweighting:** For each $i = 1..N$, simulate independent additional data $(y_{ij})_{j=J(0)+1..J(1)}$ from $p(y|\gamma^{(0)}_i)$ for $j = 1..J$. Compute the new weights:

$$w^{(1)}_i \propto \prod_{j=J(0)+1}^{J(1)} u(\gamma^{(1)}_i, y^{(1)}_{ij})$$

3. **Selection Step:** Resample $(\gamma^{(1)}_1, ..., \gamma^{(1)}_N)$ from $(\gamma^{(0)}_1, ..., \gamma^{(0)}_N)$ according to a multinomial distribution with weights $w^{(1)}_i$, note $b^{(1)}_i$ the corresponding utilities.

4. **Metropolis-Hastings Step:** For each $i = 1..N$, simulate $\gamma^{(1)}_i$ from $K_1(\gamma^{(1)}_i, .)$ and $J$ independent data sets $(y^{(1)}_{ij})$ from $p(y|\gamma^{(1)}_i)$ for $j = 1..J$. Compute $\pi^{(1)}_i = \prod_{j=1}^{J} u(\gamma^{(1)}_i, y^{(1)}_{ij})$ and the acceptance rates $\alpha_i = \min(1, (\pi^{(1)}_i K_1(\gamma^{(1)}_i, \gamma^{(1)}_{i,j})/\pi^{(1)}_i K_1(\gamma^{(1)}_i, \gamma^{(1)}_{i,j})))$. Set $\gamma^{(1)}_i = \gamma^{(1)}_i$ with probability $\alpha_i$ and $\gamma^{(1)}_i = \gamma^{(1)}_i$ with probability $1 - \alpha_i$.

We loop it along $n$.

4. Convergence issues

4.1. Overview of existing convergence theorems

We shall now focus on convergence results of our optimization algorithm. Let $\pi_\alpha$ be the probability measure over $\Gamma$ whose Lebesgue density is $U^J$ up to a normalization constant. We shall also introduce notations for some sets of test functions:

- $C_b(\Gamma)$ is the set of continuous bounded functions on $\Gamma$,
- $C_{b,1}(\Gamma)$ is the subset of $C_b(\Gamma)$ of the functions $\phi$ verifying $\|\phi\| \leq 1$ where:

$$\|\phi\| = \sup_{x \in \Gamma} |\phi(x)|$$

- $C_{b,1}^{0,\infty}(\Gamma)$ is the subset of $C_b(\Gamma)$, continuous and infinitely differentiable.
We first recall a simple convergence results shown in [1], for the two presented algorithms. For any iteration \( n \) fixed, we introduce a set of assumptions (A) as:

**Assumption 4.1. (A)** Assume that:
- \( u \) is continuous, positive and bounded,
- for all \( \gamma \) and \( n \), \( K_{1,n}(\gamma, .) > 0 \) on the support of \( h_J \),
- \( (1/N) \sum_{i=1}^{N} \text{var}_{K_{1,n}(\gamma, .)}(w_{i}^{(n)}) \) bounded independently of \( N \),
- there exists \( \delta > 0 \) such that, conditionally to formerly drawn particles \( (\gamma_{i}^{(n-1)})_{i=1..N} \):
  \[
  \left( \frac{1}{N} \sum_{i=1}^{N} \text{var}_{K_{1,n}}(w_{i}^{(n)}) \right)^{2+\delta} \sum_{i=1}^{N} E_{K_{1,n}}(w_{i}^{(n)})^{2+\delta} \longrightarrow 0, N \rightarrow +\infty
  \]
- \( \gamma \mapsto \text{var}[K_{2,n}(\gamma, .)] \) is bounded.

Assumption (A) corresponds to regularity hypothesis on Markov kernels that are easy to meet in practice. Under these conditions, one can prove the following step-by-step theorem:

**Theorem 4.1. (Step-by-Step convergence result)** Under assumption (A), for any iteration \( n \), there exists a constant \( a_n \) such that for any measurable and bounded \( \phi \) over \( \Gamma \), we have:

\[
E \left[ \left( \frac{1}{N} \sum_{i=1}^{N} \delta_{\gamma_{i}^{(n)}}(\phi) - \pi_{n}(\phi) \right)^{2} \bigg| F_{n-1} \right] \leq \frac{a_n \|\phi\|_{\infty}^{2}}{N} \tag{4.1}
\]

where \( F_{n-1} \) stands for the conditioning on the previous sample \( (\gamma_{i}^{(n-1)})_{i=1..N} \).

A detailed proof of this result can be found in [1], using Lindeberg’s types of arguments. It states that, for each iteration \( n \), the simulated sample generated by our algorithm will get closer to the searched optimum for a big enough sample size \( N \). The main limit of it is that we have no clue about how \( a_n \) might grow with \( n \), we then hope to establish an explicit dependency of \( a_n \) on \( n \). Once this obtained, it would allow us to know the number of particles we need to have a certain level of accuracy. Indeed, this level imposes a minimal value of \( J(n) \) through the convergence of the target law towards the Dirac on the mode of the utility function. Then we only have to find the optimal cooling schedule to deduce a minimal value of \( n \).

In the particular case of the Resampling-Markov (RM) algorithm, stronger results can be shown, as presented in the following theorem.

**Theorem 4.2. (Convergence of the RM algorithm)** Assuming that \( u \) is bounded by strictly positive constants, there exists a constant \( c_n \) such that for any \( \phi \) in \( C_{b}(\Gamma) \), we have:

\[
E \left[ (\pi_{n}^{N}(\phi) - \pi_{n}(\phi))^{2} \right] \leq c_n \frac{\|\phi\|_{\infty}^{2}}{N}
\]
Let $B$ be a neighborhood of the targeted optimum. Let $f_n^N(B)$ be the frequency of visiting $B$ for the RM algorithm, and $f_n(B)$ be the frequency of visiting $B$ for a theoretical trajectory from $\pi_0 \otimes \ldots \otimes \pi_n$. If the cooling schedule is linear, we have:

$$E \left[ \left( f_n^N(B) - f_n(B) \right)^4 \right]^{1/4} \leq \frac{c}{\sqrt{N}}$$

A simple proof of it is also available in [1]. Recalling that our goal is the optimum determination, this convergence of the visiting frequencies of optimum’s neighborhoods brings a valuable evidence for the effectiveness of the RM algorithm with a linear cooling schedule. We are now concerned by extending those convergence results.

4.2. The Feynman-Kac formalism

Usual central limit type of statistical theorems (used to show Theorem 4.2) are not refined enough for our purpose. That’s why we should analyze our particle systems algorithm in terms of measure processes, as pioneered by Pierre Del Moral in [12] or [13]. Our demonstration will be strongly inspired by [13] in which an uniform convergence result is brought up for sequential particle filtering. Let us introduce some useful notations related to measure theory:

On the space of signed measures on $\Gamma$, we define the total variation norm of a signed measure $\mu$:

$$\| \mu \|_{TV} = \sup \{ \mu(f), f \in C_0(\Gamma) \}$$

$M_1(\Gamma)$ is the space of all probability measures on $\Gamma$, furnished with the weak topology.

If $K$ is a Markov transition on $\Gamma$ with $B(\Gamma)$ its set of Borelians, we define the Dobrushin ergodic coefficient (see [15]) by:

$$Dob(K) = 1 - \sup_{x,z \in \Gamma} |K(x,A) - K(z,A)|, A \in B(\Gamma)$$

We can notice here that if $K$ is independent from the starting point, its Dobrushin coefficient will be 1. To interpret our algorithm in terms of discrete-time measure-valued stochastic process, we consider the following measure-valued dynamical system:

$$\pi_n = \phi_n(\pi_{n-1})$$

where $\pi_0 \in M_1(\Gamma)$ and $\phi_n : M_1(\Gamma) \to M_1(\Gamma)$ is a continuous function given by:

$$\phi_n(\pi) = \psi_n(K_{1,n}^{\otimes} \pi)K_{2,n}^{\otimes}$$

$$\psi_n(\pi)f = \frac{(1 \otimes 1)(\pi)(g_n f)}{(1 \otimes 1)(\pi)(g_n)}$$

where:
\[ g_n(\gamma, \gamma') = \frac{\prod_{j=1}^{J(n)} u(\gamma', F^{-1}(\gamma', V_j))}{K_{1,n}(\gamma, \gamma')} \]

where \( F^{-1}(\gamma', . .) \) is the vector of inverse cdf of \( p(., | \gamma') \) taken on as an independent uniform random vector \( V_j \).

By construction, \( \tilde{\pi}_n = \pi_n \otimes U_{[0,1]}^{\otimes J(n)} \) is solution of the transport equation \( \mu_n = \phi_n(\mu_{n-1}) \). It also applies to the empirical measure \( \pi_n = \frac{1}{N} \sum_{i=1}^{N} \delta_{\gamma_i} \). Indeed, recalling that \( \tilde{\gamma}_i^n \) is the state of the \( i^{th} \) particle after \( n \) iterations and the importance sampling step, and \( \gamma_i^n \) the state of the \( i^{th} \) particle after \( n \) iterations, we can write:

\[ \psi_n \frac{1}{N} \sum_{q=1}^{q=N} \delta_{\tilde{\gamma}_q^n / \gamma_1^n, ..., \gamma_N^n} = \frac{1}{N} \sum_{i=1}^{N} g_n(\tilde{\gamma}_i^n, \gamma_i^n) K_{2,n}(\tilde{\gamma}^{n+1}_i, ..., \gamma_i^n) \]

Similarly, one can write \( \tilde{\pi}_n^N \), the joint empirical measure over the augmented probability space, namely \( \frac{1}{N} \sum_{i=1}^{N} \delta_{\tilde{\gamma}_i^n, V_{i1}, ..., V_{iN}} \).

Before giving the main theorem, let us introduce the notations and relations that will help us to formulate and prove it respectively.

\[ \phi_{n/p} = \phi_n \circ \cdots \circ \phi_{p+1}, 0 \leq p \leq n. \]

with convention \( \phi_{n/n} = Id \). A simple but careful induction gives the relations:

\[ \phi_{n/p}(\pi)f = \frac{1(\otimes 1)(\pi)}{(1 \otimes 1)(\pi)} (K_{1,p} g_{n/p} (K_{2,n/p} f)) \]

where:

\[ K_{2,n/p-1}f = \frac{K_{2,p} K_{1,p} (g_{n/p} (K_{2,n/p} f))}{K_{2,p} K_{1,p} (g_{n/p})}, g_{n/p-1} = g_p K_{2,p} K_{1,p} (g_{n/p}) \]

with the conventions \( g_{n/n} = 1 \) and \( K_{2,n/n} = Id \).
Another useful notation will be to call $S_{n/p}$ the following operator:

$$S_{n/p}f = \frac{K_{2,p}K_{1,p}(g_{n/p}f)}{K_{2,p}K_{1,p}(g_{n/p})}, \quad 0 \leq p \leq n$$

so that one can easily check the following equality

$$K_{2,n/p-1} = S_{n/p}S_{n/p+1}...S_{n/n}$$

We will call

$$K_{2,p,n} = K_{2,p}K_{2,p+1}...K_{2,p+n}$$

4.3. A general convergence result

It is clear that Assumption (A) is checked by the problem we propose ourselves to solve. Indeed, a sufficient reason is that we work on truncated functions (see Section 2.4 and Section 4.5 for practical developments). So the step by step convergence theorem is true here but we are going to prove a stronger result.

The convergence theorem sought must prove the convergence of the empirical laws towards the Dirac in the mode of our utility function and establish a dependency between the number of simulations needed and the number of particles used to get a certain level of precision.

We will denote by $D$ the largest diameter over all the coordinates of $\Gamma$ and $K_{3,k} = K_{2,k}K_{1,k}$. And let $\sigma$ be such that:

$$\text{Var}(K_{3,k}) \geq \frac{\sigma^2}{J(k)}$$

In the following demonstration, we consider a Markov step which is continuous relatively to Lebesgue measure. The constants displayed are computed for Gaussian random walks, without any loss of generality. For further details, the reader can refer to section 4.5.

Our first useful result is the control on the Dobrushin coefficient of the operators $S_{p+T,k}$ as defined earlier summed up in the following

**Lemma 4.1.** For every $k$, $p$, $T$ verifying $p + 1 \leq k \leq p + T$, we get:

$$\text{Dob}(S_{p+T/k}) \geq \epsilon_k^2 \quad \text{a.s.}$$

(4.5)

with $\epsilon_k = \exp \left( -\frac{J(k)}{2} \left( \frac{mD^2}{\sigma^2} + \ln \left( \frac{U_{\text{max}}}{U_{\text{min}}} \right) \right) \right)$

**Proof.** Quick calculations with rough inequalities give us that for all borelians $A \subset \Gamma$ and for all $\gamma \in \Gamma$:

$$\epsilon_k \pi_k(A) \leq K_{3,k}(\gamma, A) \leq \frac{1}{\epsilon_k} \text{Leb}(A)$$

with $\text{Leb}$ the Lebesgue measure. Then, as

$$S_{p+T,k}f = \frac{K_{3,k}(g_{p+T/k}f)}{K_{3,k}(g_{p+T/k})}$$
with \( f \in C_{b,1}(\Gamma) \), we easily get that 
\[
S_{p+T,k}f \geq \epsilon_k^2 \frac{Leb(g_{p+T/k}f)}{Leb(g_{p+T/k})}
\]
which gives the expected result.

Now we can state the main result of this part. It is presented in the following

**Theorem 4.3. (General convergence result)** For any \( f \in C_{b,1}(E) \), \( n \in N \), \( T \leq n \), we have:
\[
E \left( \left| \pi_N^n f - \pi_n f \right| \right) \leq \frac{2T}{\sqrt{N}} \prod_{k=n-T+1}^{n} \frac{1}{\epsilon_k^2} + \prod_{k=n-T+1}^{n} (1 - \epsilon_k^2)
\]

If \( J(k) = c \ln(k) \left( \frac{mD^2}{\sigma^2} + \ln\left( \frac{U_{\max}}{U_{\min}} \right) \right)^{-1} \) with \( c \leq 1 \) and \( T \propto n^b \) with \( c < b < 1 \), this inequality gives us the dependency sought between \( n \) and \( N \) for a given level of accuracy required.

**Proof.** With the convenient notations introduced previously, the demonstration relies on the following decomposition, (telescopic sum):
\[
\tilde{\pi}_N^n f - \tilde{\pi}_n f = \sum_{p=n-T+1}^{n} \left( \phi_{n/p}(\tilde{\pi}_N^p f) - \phi_{n/p}(\phi_p(\tilde{\pi}_N^{p-1}) f) \right) + \left( \phi_{n/n-T}(\tilde{\pi}_N^{n-T}) f - \phi_{n/n-T}(\tilde{\pi}_n^{n-T}) f \right)
\]

By the triangular inequality, we get:
\[
\left| \tilde{\pi}_N^n f - \tilde{\pi}_n f \right| \leq \sum_{p=n-T+1}^{n} \left| \phi_{n/p}(\tilde{\pi}_N^p f) - \phi_{n/p}(\phi_p(\tilde{\pi}_N^{p-1}) f) \right| + \left| \phi_{n/n-T}(\tilde{\pi}_N^{n-T}) f - \phi_{n/n-T}(\tilde{\pi}_n^{n-T}) f \right|
\]

Let us notice here that we can control the \( g_n \) and the \( g_{n/p} \) corresponding to the weights involved in the selection step:
\[
C^{n-p} \beta \sum_{k=0}^{k} J(k) \leq g_{n/p} \leq C^{n-p} \alpha \sum_{k=0}^{k} J(k), \text{ a.s.}
\]

with:
\[
\alpha = U_{\max} e^{\frac{mD^2}{\sigma^2}}
\]
\[
\beta = U_{\min} e^{-\frac{mD^2}{\sigma^2}}
\]
\[
C = Vol(\Gamma)
\]

Using (4.4), we can bound a.s. each term of the sum by:
\[
\frac{b_{n/p}}{a_{n/p}} \left( \left| \tilde{\pi}_N^p f_1 - \phi_p(\tilde{\pi}_N^{p-1}) f_1 \right| + \left| \tilde{\pi}_N^p f_2 - \phi_p(\tilde{\pi}_N^{p-1}) f_2 \right| \right)
\]

with:
\[
f_1 = K_{1,p} g_{n/p} K_{2,n/p} f, \quad f_2 = K_{1,p} g_{n/p}
\]
so that \( f_1, f_2 \in C_{b,1}(\Gamma) \). This can be classically proved by introducing a third appropriate term, \( \phi_p(\tilde{\pi}_p^{N-1}) (K_{1,p}g_{n/p}K_{2,n/p}f)\tilde{\pi}_p^{N}(K_{1,p}g_{n/p})^{-1} \) and using the triangular inequality. Since \( \tilde{\pi}_p^N \) is the empirical measure associated to \( N \) independent random variables, with common law \( \phi_p(\tilde{\pi}_p^{N-1}) \), the central limit theorem gives us that:

\[
\mathbb{E} \left( |\tilde{\pi}_p^N f_1 - \phi_p(\tilde{\pi}_p^{N-1}) f_1 | \right) \leq \frac{1}{\sqrt{N}} \mathbb{E} \left( |\tilde{\pi}_p^N f_2 - \phi_p(\tilde{\pi}_p^{N-1}) f_2 | \right) \leq \frac{1}{\sqrt{N}} \tag{4.9}
\]

Summing these inequalities in (4.3), and using Jensen inequality, we get:

\[
\mathbb{E} \left( |\pi_n^N f - \pi_n f | \right) \leq \frac{2T}{\sqrt{N}} \left( \sum_{k=n-T+1}^{n} \frac{1}{\epsilon_k^2} \right) \tag{4.10}
\]

The second term is bounded (Dobrushin theorem). Indeed lemma ?? gives that:

\[
\left\| \phi_{n,n-T}(\mu) - \phi_{n,n-T}(\nu) \right\|_{TV} = \left\| \mu_{n,n-T}K_{2,n,n-T} - \nu_{n,n-T}K_{2,n,n-T} \right\|_{TV}, \ a.s. \tag{4.11}
\]

\[
\leq \prod_{k=n-T+1}^{n} (1 - \text{Dob}(S_{n/k})) \left\| \mu_{n,n-T} - \nu_{n,n-T} \right\|_{TV}, \ a.s. \tag{4.12}
\]

\[
\leq \prod_{k=n-T+1}^{n} (1 - c_k^2), \ a.s. \tag{4.13}
\]

This gives us the announced result which was the first part of the theorem. If \( J(k) = -c \ln(k) \left( \frac{\ln^2(U_{\max})}{\sigma^2} + \ln(U_{\max}/U_{\min}) \right)^{-1} \) with \( c \leq 1 \) we get \( \epsilon_k^2 = \frac{1}{c_k^2} \), so (4.10) becomes:

\[
\mathbb{E} \left( |\pi_n^N f - \pi_n f | \right) \leq \frac{2T}{\sqrt{N}} (nl)^c + e^{\frac{T}{nl}} \tag{4.14}
\]

If \( T = E(n^b) \) with \( b \) as above, we only have to take \( n \) large enough to have the second term small enough. Then, the first term gives us the adjustment needed over \( N \). This ends the proof of the theorem which proves that the algorithm of optimization is efficient.

According to the proof of this theorem, the number of particles and iterations needed are huge to have a good level of accuracy as it already appeared in the field of particle filter convergence (see [16]). Martingale arguments were used to improve the constants derived. Here, we explain how in practice these constants are not that big in Section 4.5. Before that, we show a result of uniform convergence for a simplified algorithm.
4.4. A uniform convergence theorem for the Resampling-Markov algorithm

Now we want to analyze the convergence of the particular case of the Resampling-Markov algorithm. We can notice that if we choose a linear cooling schedule for $J$, then $g_n$ are uniformly bounded, which means that one can find a constant $0 < a < 1$ such that, for any $n$:

$$a \leq g_n \leq 1/a, \ a.s.$$

Moreover, it has appeared that regularity conditions on the Markov kernel are needed to control the convergence, we will therefore assume that the Markov step $K_n$ verifies the following mixing property for each $n$: there exists $0 < \epsilon < 1$ and a probability measure $\mu_n$ such that, for all $A$ measurable:

$$\epsilon \mu_n(A) \leq K_n(x, A) \leq 1/\epsilon \mu_n(A)$$

These two last properties lead to the following uniform convergence theorem:

**Theorem 4.4. (Uniform convergence of Resampling-Markov algorithm)**

Assume that the cooling schedule is linear and that $K_n$ verifies the mixing property with $\epsilon = 1 - \exp(-\theta)$. Then for any $f \in C_b, 1(\Gamma)$ :

$$\sup_{n \in \mathbb{N}} E \left( |\pi_n f - \pi_n f| \right) \leq \frac{C}{N^\alpha}$$

with $C = \frac{5}{\sigma^2}$ and $\alpha \geq \frac{\theta}{2\sigma - 2 \log a} > 0$.

**Proof.** The proof of this result is an obvious particular case of the proof of theorem 2, with $\epsilon_k = \epsilon$ in lemma 1. The same calculations lead us to:

$$\sup_{n \in \mathbb{N}} E \left[ |\pi_n f - \pi_n f| \right] \leq \frac{4T}{\sqrt{N}} \frac{1}{a^T} + (1 - \epsilon)^T$$

Choosing the linking function:

$$T(N) = 1 + \left[ \frac{1}{4} \frac{- \log N}{\log a (1 - \epsilon)} \right]$$

where $[\cdot]$ stands for the integer part, it comes:

$$\sup_{n \in \mathbb{N}} E \left[ |\pi_n f - \pi_n f| \right] \leq \frac{C}{N^\alpha}$$

with $C$ and $\alpha$ as defined in theorem.

4.5. A practical point of view

In this section, we want to precise the choices of jump functions, according to convergence aspects presented in the previous section, and we discuss hypothesis used to prove our main convergence theorem.
4.5.1. A practical choice for the importance sampling proposal

As formerly said, a major concern is the choice for the range of the variance of the transition functions over $\Gamma$. Let us first analyze how one can propose a smart choice for the variance of the importance sampling proposal $K_{1,n}$. For this purpose, we should recall the normalized weights at iteration $n$:

$$w_i^{(n)} = \frac{\prod_{j=1}^{J} u(\tilde{\gamma}_i^{(n)}, \tilde{\gamma}_j^{(n)})/K_{1,n}(\tilde{\gamma}_i^{(n-1)}, \tilde{\gamma}_i^{(n)})}{\sum_{i=1}^{N} \prod_{j=1}^{J} u(\tilde{\gamma}_i^{(n)}, \tilde{\gamma}_j^{(n)})/K_{1,n}(\tilde{\gamma}_i^{(n-1)}, \tilde{\gamma}_i^{(n)})}$$

Considering that $N$ and $J$ are big enough to allow Large Number Law’s type of approximations, it comes:

$$\log(w_i^{(n)}) = J \left( \frac{1}{J} \sum_{j=1}^{J} \log(u(\tilde{\gamma}_i^{(n)}, \tilde{\gamma}_j^{(n)})) \right) - \log(K_{1,n}(\tilde{\gamma}_i^{(n-1)}, \tilde{\gamma}_i^{(n)})) \quad (4.15)$$

and

$$-\log(\frac{N}{J} \sum_{i=1}^{N} \prod_{j=1}^{J} u(\tilde{\gamma}_i^{(n)}, \tilde{\gamma}_j^{(n)})/K_{1,n}(\tilde{\gamma}_i^{(n-1)}, \tilde{\gamma}_i^{(n)})) - \log(N) \quad (4.16)$$

Then, the following approximation of the weights holds:

$$w_i^{(n)} = \frac{\exp \left( J I(\tilde{\gamma}_i^{(n)}) \right) U^J(\tilde{\gamma}_i^{(n)})}{NK_{1,n}(\tilde{\gamma}_i^{(n-1)}, \tilde{\gamma}_i^{(n)}) \int U^J(\gamma) d\gamma} (1 + O(\frac{1}{\sqrt{J}})) \quad (4.17)$$

where $I(\gamma) = E_{y|\gamma}[\log(u(\gamma, y)/U(\gamma))]$.

From equation (4.17), we can first notice that:

$$\mathbb{E}_{K_{1,n}(\tilde{\gamma}_i^{(n-1)}, \cdot)}[w_i^{(n)}] = \left( \frac{1}{N} \int \exp(JI(\gamma)) d\pi_n(\gamma) \right) (1 + O(\frac{1}{\sqrt{J}}))$$

So, for our order of approximation, we get that:

$$\int \exp(JI(\gamma)) d\pi_n(\gamma) = 1 + O(\frac{1}{\sqrt{J}}) \quad (4.18)$$

Moreover, assuming that $K_{1,n}$ is chosen as a Gaussian random walk with variance $\sigma_{1,n}^2$, and using (4.17), it comes:

$$w_i^{(n)} = \left( \frac{\sqrt{2\pi \sigma_{1,n}}}{N} \exp \left( J I(\tilde{\gamma}_i^{(n)}) + \frac{1}{2} X^2 \right) \int U^J(\tilde{\gamma}_i^{(n)}) d\gamma \right) \left( 1 + O(\frac{1}{\sqrt{J}}) \right) \quad (4.19)$$

where $X \sim N(0,1)$. Considering we need $\mathbb{E}_{K_{1,n}(\tilde{\gamma}_i^{(n-1)}, \cdot)}[w_i^{(n)}] = 1/N$ and that $\exp(\sqrt{\text{var}(X)})/2 = \left(1 + O(\frac{1}{\sqrt{J}})\right)$, we get an equation that should be verified by $\sigma_{1,n}$:

$$\sigma_{1,n} = \left( e^{1/2} \sqrt{2\pi} \mathbb{E}_{K_{1,n}(\tilde{\gamma}_i^{(n-1)}, \cdot)}[\exp(JI(\gamma)) \int U^J(\gamma) d\gamma] \right) \left( 1 + O(\frac{1}{\sqrt{J}}) \right) \quad (4.20)$$
We want now to get an approximation of 
\[ \frac{U_J(\tilde{\gamma}(n))}{\int U_J(\gamma) d\gamma}. \]
Assuming that the points \( \tilde{\gamma}(n) \) are relatively close to the optimum (when \( J \) is “large” enough, as supposed), we are allowed to use a Taylor development of \( U \) around its mode:
\[
U_J(\tilde{\gamma}(n)) = \left( U^* - \frac{1}{2} H^*(\tilde{\gamma}(n) - \gamma^*)^2 + o(\sigma^2_{1,n}) \right)^J
\]
(4.21)

using:
\[
\left( 1 - \frac{H^*(\tilde{\gamma}(n) - \gamma^*)^2}{2U^*} \right)^J = \exp \left( -\frac{JH^*(\tilde{\gamma}(n) - \gamma^*)^2}{2U^*} \right) (1 + o(\sigma^2_{1,n}))
\]
it comes:
\[
\int U_J(\gamma) d\gamma = \sqrt{\frac{2\pi U^*}{JH^*}} (1 + o(\sigma^2_{1,n}))
\]
(4.22)

In addition, as we want to reach \( \gamma^* \) from \( \tilde{\gamma}(n) \) by \( K_{1,n} \), we need \( (\tilde{\gamma}(n) - \gamma^*)^2 = \sigma^2_{1,n} + o(\sigma^2_{1,n}) \), so, using (4.21) and (4.22), we can write:
\[
\frac{U_J(\tilde{\gamma}(n))}{\int U_J(\gamma) d\gamma} = \sqrt{\frac{2\pi U^*}{JH^*}} \exp \left( -\frac{JH^*(\tilde{\gamma}(n) - \gamma^*)^2}{2U^*} \sigma^2_{1,n} \right) (1 + o(\sigma^2_{1,n}))
\]

This last approximation represents the fact that, for a sufficiently large \( J \), the density of \( \pi_n \) can be locally approximated around the mode \( \gamma^* \) by a Gaussian distribution with mean \( \gamma^* \) and with variance \( \frac{U^*}{JH^*} \). Then, injecting this approximation in (4.20), we can easily check that:
\[
\sigma_{1,n} = \sqrt{\frac{U^*}{JH^*}}
\]
(4.23)
is a solution of (4.20), noticing that, with this choice of \( \sigma_{1,n} \) and using (4.18):
\[
E_{K_{1,n}(\gamma_{n-1},,)} [\exp(JI(\gamma))] = \int \exp(JI(\gamma))d\pi_n(\gamma) + O \left( \frac{1}{\sqrt{J}} \right)
\]
(4.24)
\[
= 1 + O \left( \frac{1}{\sqrt{J}} \right)
\]
(4.25)

Following the same computations, it is easy to show that we can generalize this result for \( \text{dim}(\Gamma) > 1 \) choosing:
\[
\sigma_{1,n} = \sqrt{\det(\frac{U^*}{J}H^{*-1})}
\]
(4.26)
where \( H^* \) is now the Hessian matrix of \( U \) on the optimum.
4.5.2. A practical point of view on the Markov renewal step

As we saw in the convergence analysis, a key property of the Markov kernels for fast convergence is the “mixing property”, which can be typically met when $K_{2,n}$ has a continuous Lebesgue derivative, as supposed in the demonstration.

However, as we introduced it, a convenient choice to build such Markov steps adequately to the right target distribution is the Metropolis-Hastings kernel, which is definitely not Lebesgue-derivable. A practical and heuristic conclusion from this may be: the higher is the acceptance probability in the Metropolis-Hastings step, the faster is the convergence of the algorithm. As a consequence, using an adaptive Markov step (in which this acceptance rate is made to be larger, see [23]) could significantly improve the convergence. More case-specific Markov step could be used, such as Gibbs kernels which are then Lebesgue-derivable.

As for the variance of this Markov proposal, we choose it lower bounded. Combined with the choice of the variance of the importance sampling proposal (cf 4.26), this would roughly ensure the variance assumption (4.3). It also allows us a larger scale of exploration of $\Gamma$, as desired. Note that this condition on the variance proposal does not imply any bounding condition on the resulting Metropolis-Hastings kernel.

4.5.3. Empirical remarks on the convergence in practice

We end this practical point of view by considerations on the effective convergence speed in practice. Indeed, the constant involved in the upper bound of theorem ?? can be very large in all generality, as noticed in a particle filter environment in [16]. As a matter of fact, the determination of this constant was made by rough inequalities which could be obviously improved (as managed in [16] by Jacod and Del Moral by martingale arguments). Indeed, the total variation distance between the empirical and the target distributions was roughly bounded by 1 although it should empirically tends towards 0 with iterations. More subtle bounds would be found by linking this distance to iterations $n$, but this should be either mathematically more complicated, or less general (only for specific kernels). Another consequence of this fact is that the utility function could be adapted by proper dynamic truncatures when samples get closer to the mode as iterations increase, in order to reduce the ratio $U_{\text{max}}/U_{\text{min}}$ (which obviously improves the convergence).

Let us conclude with the following remark: using the logarithmic cooling schedule of theorem ?? would require $\left(\frac{mD^2}{\sigma^2} + \ln\frac{U_{\text{max}}}{U_{\text{min}}}\right)$ to be small in order to quickly reach the final $J$ determined by the wanted level of accuracy. However, this is equivalent to reduce the compact size and/or the ratio $U_{\text{max}}/U_{\text{min}}$, driving to an over-simplified optimization problem or to a hopeless algorithm. Let us recall that the constants displayed are not optimal, but they give an intuition of the different variables impacts.
5. Reliability of our joint calibration challenge

In this section, we present practical applications of the particle algorithm to the introductory joint calibration problem, in order to test on simple examples the effectiveness of the method. For this purpose, we calibrate the historical and risk-neutral parameters of two affine pricing models: Black-Scholes and a latent stochastic volatility model. We name “estimation” the determination of the parameters as explained earlier, within a decision analysis framework. It does not refer to the estimation in the statistical sense, and confidence intervals do not really make sense in such a context.

5.1. Implementation issues

We want here to describe the technical and numerical choices made for the various applications presented. First, the necessary algorithmic choices are detailed. For more generality, we used the general algorithm in the simulations. Two Markov kernels $K_{1,n}$ and $K_{2,n}$ are then required, as well as the number $N$ of particles and the cooling schedule $J(n)$. As Markov proposals, we simply choose truncated normal random walks as evoked in 4.5. As for sample size $N$, central limit type inequalities would lead us to typical values of order $N = 10,000$ particles to give accurate Monte-Carlo integral approximations. However, as we are not interested by integral approximations but only by the mode of target distributions, this number might be too large for our purpose. In practice, it appears that $N$ below 1,000 gives good results in most cases. In our simulations, we take $N = 500$. Note that this sample size should obviously be larger for the RM algorithm, since in practice, it suffers more from degeneracy. As cooling schedule, our theoretical arguments suggest a logarithmic form for small values of $J$, and a linear form for large ones (when the algorithm is almost “reduced” to a RM algorithm). For computational convenience, we simply take a linear cooling scheme from 1 to 31 by a time-step of 2.

In the computation of the objective function $L$ defined in Section 2.4, we choose weights of the form:

$$\omega_i^{-1} = \text{Vega}(T_i, K_i)$$

where $\text{Vega}(\cdot)$ is the Black-Scholes vega of the option computed using the market implied volatility. The interest of this weight, as noted by Cont-Tankov ([11]), is that it converts errors in price in errors in implied volatility, thus rescaling all terms entering in $L$ to the same order of magnitude.

For each day of our sample, we need to produce an estimate of the spot volatility level. A practical way to obtain this was to evaluate this volatility from an historical analysis. The statistical estimate used was a moving empirical average variance in order to capture day-to-day effects and both persistence and clustering of volatility.

Regarding the entropy term, it caused numerical problems linked to the choice of the constant added to make $U$ positive. Indeed, after 3 iterations, the entropy term explodes and the calculation of the coefficient $\alpha$ gets hazardous so we could
not use in a satisfying way the advantages of the entropy. However, the expected convexification effect was actually observed these 3 first iterations.

5.2. Application to a simulated data set

5.2.1. Simulation using Black-Scholes model

In the first simulation study, we generated a panel of stocks and option prices using the Black-Scholes model. One main advantage of this specification is its simplicity which allows to precisely test the cross-sectional effect. Following market price of risk specification enlightened in Section 2, we fix $P$ and $Q$ parameters respectively to $(\mu; \sigma) = (0.08; 0.4)$ and $(\mu = r; \sigma) = (0.037; 0.4)$. A 30-days long returns series was simulated under $P$-parameters and option prices were computed under the measure $Q$ using the Black-Scholes formula. The maturity of theses options was one year and we used 21 equidistant strikes from 30 to 50 for a spot being at 40.

Our approach to capture cross-sectional effect is made to produce independent estimations with different stocks returns length, beginning from a short horizon of 5 days and iteratively by a time-step of size 5 reaching a 30 days long horizon. We estimate volatility parameter $\sigma$, drift under the historical measure $\mu^P$ and market price of risk $\lambda$. The estimated parameters are shown in next table. We reported optimal parameter estimates that is to say the value corresponding to the minimum of our utility functional (overall option fit). LSE line gives the Least Square Error incurred.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$T = 5$</th>
<th>$T = 10$</th>
<th>$T = 15$</th>
<th>$T = 20$</th>
<th>$T = 25$</th>
<th>$T = 30$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu^P$ (0.08)</td>
<td>0.068</td>
<td>0.074</td>
<td>0.086</td>
<td>0.072</td>
<td>0.073</td>
<td>0.088</td>
</tr>
<tr>
<td>$\sigma^P,Q$ (0.4)</td>
<td>0.344</td>
<td>0.355</td>
<td>0.372</td>
<td>0.391</td>
<td>0.429</td>
<td>0.433</td>
</tr>
<tr>
<td>$\lambda$ (0.1)</td>
<td>0.073</td>
<td>0.076</td>
<td>0.144</td>
<td>0.102</td>
<td>0.097</td>
<td>0.129</td>
</tr>
<tr>
<td>$LSE$</td>
<td>0.58</td>
<td>1.03</td>
<td>1.75</td>
<td>1.85</td>
<td>2.13</td>
<td>2.81</td>
</tr>
</tbody>
</table>

A number of broad points emerge from Table 1. Longer time horizons provide significant pricing improvement than shorter horizon, as the $LSE$’s of the pricing errors are relatively stable even though the parameters from time $T = 5$ to time $T = 30$ days are progressively larger and more difficult to identify. Once again, let us recall that we constrain the risk-neutral parameters to be time-consistent with the objective measure dynamics and we use both stocks and option prices spanning a long time period so that our results are not driven by a specific episode.

To assess the double quality (both on the historic side and on the risk neutral one) of our particle algorithm, the final sample can be plotted, cf. Figure 2 and in
the same way the least squares criteria between simulated and market option prices.

![Simulated Particles Distributions](image)

**Fig. 2.** Distributions of Simulated Particles.

We also illustrate, in this particular case, the interest of the simulated annealing. In the Figure 3, one can observe the convexification/discrimination effects with the evolution of the scales.

### 5.2.2. Simulations using Heston model

In the second series of simulation, our goal is two-fold. First, we examine how risk premiums affect option prices and then the efficiency of our algorithm to recover the historical latent stochastic volatility. The analysis has been conducted in the Heston framework, a model widely used by practitioners. Stock returns and volatility paths were generated under a specified historical measure $\mathbb{P}$, while option prices were computed under an appropriate measure $\mathbb{Q}$ for 21 equidistant strike values. The diffusion can be written as:

\[
\begin{align*}
dS_t &= \mu S_t dt + S_t \sqrt{V_t} dW^s_t - S_t \bar{\mu} \lambda dt \\
dV_t &= \kappa_v (\theta_v - V_t) dt + \sigma_v \sqrt{V_t} dW^v_t
\end{align*}
\]  

where $W^k_t$, $k = (s,v)$ are two Brownian motions with $E[dW^s_t dW^v_t] = \rho dt$. As Eraker, we chose the following shape of market prices of risk:

- return market price of risk:

\[
\lambda^S(t) = \frac{\mu^P - \mu^Q}{\sqrt{V(t)}}
\]
volatility market price of risk:

$$\lambda^V(t) = -\frac{\kappa^P - \kappa^Q}{\sigma_v \sqrt{1 - \rho^2}}$$

We led the same estimation strategy that for Black-Scholes model and produced separate estimates for time horizon ranging from $T = 5$ to $T = 30$. The estimated parameters for both measures are reported in Table 2.

Once again, the results are satisfying and interesting. Indeed, without using a huge number of simulations and particles, a good level of accuracy has been reached so it is far from being absurd to rely on the results produced by our procedure. Though, we can notice that the joint-calibration is more difficult for long time horizon data which was predictable because of the more complex and numerous effects to catch. It was not true with Black Scholes: it can be explained by the difficulty entailed by the diffusion of the volatility. But in practice, this procedure will not be of any interest for horizons larger than 10 days. Besides, a kind of stability over time seems to emerge from the results. This is an important point to examine. This could confirm one of the motivations of the resolution of this problem.

Once these points have been observed, we are naturally led to apply the algorithm to real data. We try and exploit it as much as possible notably regarding the information extracted from the use of stock returns, contained in the risk premiums.
6. Risk premiums investigation from a real-world data set

In order to assess the usefulness of our joint calibration approach, we address two option pricing issues: selecting the appropriate model and quantifying the risk premiums of the various underlying factors. In this attempt, we used the information from the cross-section of EuroStoxx 50 stocks and options series for the period June 2003 to April 2004.

6.1. Models specification

We consider the general framework displayed in Section 2 and study models incorporating stochastic volatility and jumps in both returns and volatility. On the probability space \((\Omega, \mathcal{F}, P)\), the equity price index \(S_t\) and its spot variance \(V_t\) are assumed to jointly verify:

\[
\begin{align*}
\frac{dS_t}{S_t} &= \mu S_t dt + \sqrt{V_t} dW^s_t + d \left( \sum_{n=1}^{N_t} S_{\tau_n} \left[ e^{Z^s_n} - 1 \right] \right) - S_t \bar{\mu} \lambda dt \\
\frac{dV_t}{V_t} &= \kappa v (\theta_v - V_t) dt + \sigma_v \sqrt{V_t} dW^v_t + d \left( \sum_{n=1}^{N_t} Z^v_n \right) 
\end{align*}
\]

where \(W^k_t, k = (s, v)\) are two Brownian motions with \(E[dW^s_t dW^v_t] = \rho dt\) and \(N_t\) is a Poisson process with intensity \(\lambda\) independent of the two diffusion processes. \(Z^s_n, Z^v_n \sim N(\mu_s + \rho_s Z^v_n, \sigma_s^2)\) are the jumps in returns and \(Z^v_n \sim \exp(\mu_v)\) are the jumps in volatility and \(\bar{\mu}_s = \exp \left( \mu_s + \frac{1}{2} (\sigma_s)^2 \right)\).

From this very general formulation, we derive four specifications: SV and SVJ models assume that there are respectively no jumps at all and no jumps in volatility,
SVIJ and SVCJ models allow both types of jumps and consider respectively that jumps sizes in returns are independent, correlated (parameter $\rho_s$) with those in volatility.

The market generated by the structure (6.1) and (6.2) is incomplete. Therefore, there is a multiplicity of equivalent martingale measures, corresponding to the absence of arbitrage. For the highest flexibility of the equivalent martingale measure, we assume a very general change of measures. According to the generalized Girsanov theorem recalled in Section 2, the measure transformation for Brownian motions only shifts the drift of the stochastic differential equations, while measure transformation for jump processes are more flexible. As we choose a specification with constant intensity and time-independent jump sizes, we only require that the two distributions are absolutely continuous with respect to the other one. Finally, under the risk-neutral probability measure $Q$, the equity index and its variance verify:

$$dS_t = rS_t dt + S_t \sqrt{V_t} dW^s_t (Q) + d \left( \sum_{n=1}^{N_t(Q)} S_{\tau_n} \left[ e^{Z_n(Q)} - 1 \right] \right) - S_t \mu^Q_s \lambda dt$$

$$dV_t = \kappa^Q \left( \theta^Q - V_t \right) dt + \sigma^Q \sqrt{V_t} dW^v_t (Q) + d \left( \sum_{n=1}^{N_t(Q)} Z^v_n (Q) \right)$$

where $\mu^Q_s = \exp \left( \mu^Q_s + \frac{1}{2} (\sigma^Q_s)^2 \right)$. We define the sets of structural $\gamma^P = (\mu, \kappa_v, \theta_v, \lambda, \mu_s, \rho_s, \sigma_s, \mu_v)$ and implicit $\gamma^Q = (\kappa^Q, \theta^Q, \lambda^Q, \mu^Q_s, \rho^Q_s, \sigma^Q_s, \mu^Q_v)$ parameters. We also note that the change of measures constrains $\sigma_v$ and $\rho$ to be the same under both measures.

6.2. Parameter determination using cross-section

The presented methodology has been applied to the cross-section of stock returns and option prices. Nevertheless, the extreme computational burden generated when using both sources of data severely constrained how much and what type of data could be used. Therefore, we focused on three sub-periods of different horizons of our data set and discussed the option pricing implications of our results. These three Eurostoxx 50 scenarios are denoted by $S_i$, $i = (1, 2, 3)$.

6.2.1. Parameter estimates

Table 3 reports parameter estimates under both measures $P$ and $Q$ for the SV and SVJ models and three scenarios. The first scenario lasts from 02/06/03 to 06/06/03, the second from 09/06/03 to 13/06/03 and the third one is the joint of the two preceding, spanning from 02/06/03 to 13/06/03. Maturity of all the scenarios is 183 days. The estimates are quoted in annualized form in order to be comparable with existing results in option pricing literature (e.g. [4] and [31]).
Table 3. Joint $P$ and $Q$ parameters estimates.

<table>
<thead>
<tr>
<th></th>
<th>SV</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$S_1$</td>
<td>$S_2$</td>
<td>$S_3$</td>
<td></td>
<td>$S_1$</td>
<td>$S_2$</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.072</td>
<td>0.078</td>
<td>0.069</td>
<td>0.069</td>
<td>0.075</td>
<td>0.064</td>
</tr>
<tr>
<td>$\kappa_v$</td>
<td>7.009</td>
<td>6.874</td>
<td>7.109</td>
<td>7.127</td>
<td>6.938</td>
<td>7.261</td>
</tr>
<tr>
<td>$\theta_v$</td>
<td>0.028</td>
<td>0.027</td>
<td>0.023</td>
<td>0.028</td>
<td>0.026</td>
<td>0.022</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>0.06</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td>$\mu_s$ (%)</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>-3.75</td>
<td>-4.01</td>
<td>-4.66</td>
</tr>
<tr>
<td>$\sigma_s$ (%)</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>4.07</td>
<td>3.17</td>
<td>6.63</td>
</tr>
<tr>
<td>$\sigma^Q_{\mu}$</td>
<td>0.541</td>
<td>0.558</td>
<td>0.472</td>
<td>0.542</td>
<td>0.612</td>
<td>0.529</td>
</tr>
<tr>
<td>$\rho^Q_{\mu}$</td>
<td>-0.306</td>
<td>-0.381</td>
<td>-0.540</td>
<td>-0.379</td>
<td>-0.431</td>
<td>-0.582</td>
</tr>
<tr>
<td>$\mu^Q = r$</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>$\kappa^Q_v$</td>
<td>4.521</td>
<td>4.756</td>
<td>4.781</td>
<td>4.368</td>
<td>4.679</td>
<td>4.804</td>
</tr>
<tr>
<td>$\theta^Q_v$</td>
<td>0.031</td>
<td>0.024</td>
<td>0.022</td>
<td>0.028</td>
<td>0.025</td>
<td>0.026</td>
</tr>
<tr>
<td>$\lambda^Q$</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>$\mu^Q_s$ (%)</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>-2.59</td>
<td>-2.89</td>
<td>-3.23</td>
</tr>
<tr>
<td>$\sigma^Q_s$ (%)</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>2.89</td>
<td>2.21</td>
<td>4.91</td>
</tr>
</tbody>
</table>

There are several interesting features to point out from these parameter estimates. First, we note that for a specified model, reported values are relatively stable across scenarios. Moreover, the joint scenario $S_3$ produced estimates close to those obtained in sub-samples $S_1$ and $S_2$. These results give evidence of the stabilization effect of our procedure in estimating model parameters across time. By incorporating both stocks and options cross-sectional aspects in the calibration, our joint approach may capture trend effects and produces also a sort of benchmark model.

Second, the difference between estimates under the two measures (historical minus risk-neutral) is the risk premium associated with Brownian, volatility or jump risk. All of these premiums were estimated to be positive across all models and scenarios. This implies that investors are averse to changes in Brownian, volatility or jump dimensions. This remark is of particular interest from the perspective of option prices: when the market is more (resp. less) volatile or jumpy, the options are more (resp. less) expensive than those implied by the objective measure as investors require higher (resp. lower) premiums.

### 6.2.2. Market smiles fit

Having estimated the $P$ and $Q$ parameters, we now discuss the empirical performance of the various models in fitting the historical implied volatility smiles. Figure 4 plots the model and market Black-Scholes implied volatilities (IVs) and presents evidence on the fit of the SV and SVJ models. We can notice that the IVs curves are very similar from one model to another and on average they fit quite well the
The upper left and right graphs in Figure 4 superimpose the five daily IVs curves for respectively scenarios $S_1$ and $S_2$. As moneyness is different from one day to another, we represent volatility curves with respect to options instead of moneyness, so that to each X-axis point corresponds an option whose daily implied volatilities are reported along the Y-axis ranging from the first to the last day. Besides, for each day, market curve goes through the big points, while model smile is represented by a solid line. We remark that for each day and scenario, SV or SVJ model smiles fit relatively well market values. The lower graphs present how the presented methodology allows to reconstruct term-structure of implied volatilities across options and time. While the reconstructed volatility surface (right) is more perturbed than the original model surface (left), the error is reasonably small.

### 6.3. Risk premium estimates and option prices effects

Our joint calibration approach attempts to simultaneously capture the effect of both the historical and implied probability measures in a one-stage procedure. Since the methodology requires the absence of arbitrage, $\sigma_v$ and $\rho$ should be the same under $P$ and $Q$. Therefore, the estimation of risk premium parameters is central to our methodology. A thorough analysis of risk premium estimates and effects is also of interest.
6.3.1. *Inference on risk premiums*

To see how risk premiums affect conditional moments of returns and volatility, Table 4 provides the instantaneous first and second moments of $Y_t = \ln(S_t)$ and $V_t$ for the SV and SVJ models.

For example, risk premiums affect both the level and mean-reversion of volatility, which implies that a positive volatility risk premium generates a structural volatility higher than its risk-neutral counterpart. Besides, jumps in returns and volatility generate different patterns of conditional non-normalities. Jumps in returns result in decreasing amounts of excess skewness and kurtosis while jumps in volatility provide a factor that combines features from both jumps in volatility and diffusive stochastic volatility.

<table>
<thead>
<tr>
<th>$\mathbb{E}[Y_t]$</th>
<th>$\mu - 0.5\mathbb{E}[V_t]$</th>
<th>$\mu - 0.5\mathbb{E}[V_t] - \lambda(\mu - \mu_s)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Var}[Y_t]$</td>
<td>$V_t$</td>
<td>$V_t + \lambda(\mu_s^2 + \sigma_s^2)$</td>
</tr>
<tr>
<td>$\mathbb{E}[V_t]$</td>
<td>$\theta_v$</td>
<td>$\theta_v + (\kappa_v)^{-1}\lambda\mu_v$</td>
</tr>
<tr>
<td>$\text{Var}[V_t]$</td>
<td>$\sigma^2 V_t$</td>
<td>$\sigma^2 V_t$</td>
</tr>
</tbody>
</table>

An analysis of conditional moments provides a theoretical description of why risk premiums are difficult to estimate. In fact, variations in these parameters slightly affect the first moment of $Y_t$ and $V_t$ and should have a small effect on the cross-section of returns and options series. A satisfactory identification of these parameters would thus require derivative contracts which solely depend on how quickly the conditional moments of the state variables $Y_t$ and $V_t$ fluctuate.

To assess further the estimation challenge of risk premiums, we illustrate in a smiles variation study why such parameters are difficult to identify. Figure 5 shows how variations of risk premiums affect implied volatility.

6.3.2. *From $\mathbb{P}$ to $\mathbb{Q}$ implied volatility*

The effects of risk premiums can also be assessed on the basis of option prices. Figure 6 displays Black-Scholes implied volatility curves for the SV and SVJ models for two maturities and prices computed under $\mathbb{P}$ and $\mathbb{Q}$.

The first smile is based on $\mathbb{Q}$ parameters while the second one is based on $\mathbb{P}$ parameters which include the effects of risk premiums estimates. One can notice that the two models generated quite similar implied volatility curves. Besides, spreads between the two measures are more severe as maturity increases and assess for uncertainty and risk averse behaviour of investors. Therefore, given a sufficient number of parameters and allowing them to change from one measure to another without constrains, one cannot distinguish different models in an analysis only based...
Joint Calibration of Option Pricing Models via Particle Methods

Fig. 5. Effects of variations of volatility and jump risk premiums on implied volatility curves for two maturities in the SV and SVJ models.

Fig. 6. Implied volatilities curves for SV and SVJ models based on parameters from $P$ and $Q$.

on option prices. And quite naturally, a simpler model will always be preferred. From a classical calibration perspective where ease and reliability are of first importance,
this is not a problem. However, if one wishes to jointly fit both cross-section of returns and options series and to determine a relevant benchmark model, another strategy has to be set up. As a consequence, our joint calibration procedure might help to select the best model(s) to take into account all aspects of the data set.

6.4. Time inhomogeneity and aversion to model misspecification

Given our joint calibration procedure, we next investigate the time homogeneous property of the models specification and the behaviour of investors towards Brownian, volatility and jump risk, described by risk premiums, across time. To answer this question in a satisfactory way, we set up a procedure which consists of estimating risk premiums parameters for SV model for each month of our data set. The Table 5 reports relevant parameter estimates.

An important conclusion emerges from this result. The divergence between the information embedded in the measures $\mathbb{P}$ and $\mathbb{Q}$ is indicative of time-varying risk premiums. More interestingly, returns and options series displayed opposite effect in terms of moments properties as measured by the skewness and kurtosis: while returns displayed negative skewness and slight kurtosis, options are characterized by strong and positive values of these moments. This effect could be related to investor’s anticipations towards model misspecification which lead to overprice Brownian, volatility and jump risk by putting high risk premiums. In fact, if the models are misspecified along certain dimensions and agents are both risk and uncertainty-averse to this model failure, then the historical and implied distributions present a form of time-varying aversion. Here again, in a classical calibration perspective, this aspect is not crucial as practitioners preclude model indeterminacy by frequently
rebalancing model parameters. On the contrary, our joint and cross-sections calibration procedure may benefit from studying models that allow time-varying parameters in their specification and under this setup might help to investigate time inhomogeneity of the historical and implied distributions.

7. Conclusion

We developed an innovative and practicable methodology for a joint calibration challenge: from a time series of stock returns and option prices, we managed to determine the characteristic parameters of both historical and risk-neutral measures. This was made possible by a sophisticated particle sampler where usual methods would have failed. Our theoretical study of it led us to efficiently tune the algorithmic preferences.

The interest of this approach is two-fold: first, it stands for a quest for information embedded in real data, complementary to the one led through classical calibration (embedded in the stochastic model). It operates a kind of average of this calibration and might be used as a benchmark. Moreover, it takes advantage of more information which could be notably used to calibrate parameters usually determined by rough calculations or intuition or a simple historical study. For instance, the idea briefly alluded to in the introduction remains a good one: the determination of unobservable parameters like correlations in the multi-dimensional case is often achieved through historical statistics independently from other parameters. With this algorithm, it could be done in a more sophisticated way and since the at-stake of these correlation products is tremendous, it is not a pointless remark.

The second interest is directly linked to the theoretical and more general information brought by the joint calibration. In particular, the last study on risk premiums is quite precious, and must be deepened. In fact, our work sheds light on the relationship between the investors and the model used: leading this study through different periods of time could be very enriching regarding our understanding of the models, their limitations and the trust investors have in them.

Then, this work gives way to a lot of interesting tracks to be explored. The first one is to optimize the conception of the algorithm. It would be very helpful to try a more dynamical optimization with less rough approximations as suggested in Section 4.5. It could allow us to establish a clear and refined dependency between the cooling schedule, the shape of the utility function and the size of the compact considered. Besides we could exploit in a better way the impact of entropy with a dynamic determination of the constant imposing the utility to be positive. It would prevent it from exploding in the calculation of the entropy and so the entropy could bring its convexification effect as suggested by our first attempts. After these improvements the convergence could be faster and the algorithm could give quicker accurate results.
8. Acknowledgement

We would like to thank Vincent Lacoste, Eric Parent and Henri Berestycki for a careful reading of this work. We are also grateful to Christian Robert and Frédéric Bois for helpful support on the elaboration of the algorithm, and to Nicolas Chopin for comments on it. At last, we thank Rama Cont, Rémi Bourrette, Yoann Bourgeois and Laurent Sifaoui for their advice and comments about the financial problem.

References

2003.


