HAL open science

Modélisation de la dépendance et simulation de processus en finance

Mohamed Sbai

To cite this version:

Mohamed Sbai. Modélisation de la dépendance et simulation de processus en finance. Mathématiques [math]. Université Paris-Est, 2009. Français. NNT : . tel-00451008v1

HAL Id: tel-00451008

https://theses.hal.science/tel-00451008v1

Submitted on 27 Jan 2010 (v1), last revised 24 Nov 2010 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse présentée pour l'obtention du titre de

Docteur de l'Université Paris-Est

Spécialité : Mathématiques appliquées

par

Mohamed SBAI

Modélisation de la dépendance et simulation de processus en finance

Thèse soutenue le 25 novembre 2009 devant le jury :

Vlad BALLY	Examinateur
Jean-David FERMANIAN	Examinateur
Emmanuel GOBET	Rapporteur
Benjamin JOURDAIN	Directeur de thèse
Antoine LEJAY	Rapporteur
Francesco RUSSO	Président du jury

Résumé

La première partie de cette thèse est consacrée aux méthodes numériques pour la simulation de processus aléatoires définis par des équations différentielles stochastiques (EDS). Nous commençons par l'étude de l'algorithme de Beskos *et al.* [13] qui permet de simuler exactement les trajectoires d'un processus solution d'une EDS en dimension 1. Nous en proposons une extension à des fins de calcul exact d'espérances et nous étudions l'application de ces idées à l'évaluation du prix d'options asiatiques dans le modèle de Black & Scholes. Nous nous intéressons ensuite aux schémas numériques. Dans le deuxième chapitre, nous proposons deux schémas de discrétisation pour une famille de modèles à volatilité stochastique et nous en étudions les propriétés de convergence. Le premier schéma est adapté à l'évaluation du prix d'options *path-dependent* et le deuxième aux options *vanilles*. Nous étudions également le cas particulier où le processus qui dirige la volatilité est un processus d'Ornstein-Uhlenbeck et nous exhibons un schéma de discrétisation qui possède de meilleures propriétés de convergence. Enfin, dans le troisième chapitre, il est question de la convergence faible trajectorielle du schéma d'Euler. Nous apportons un début de réponse en contrôlant la distance de Wasserstein entre les marginales du processus solution et du schéma d'Euler, uniformément en temps.

La deuxième partie de la thèse porte sur la modélisation de la dépendance en finance et ce à travers deux problématiques distinctes : la modélisation jointe entre un indice boursier et les actions qui le composent et la gestion du risque de défaut dans les portefeuilles de crédit. Dans le quatrième chapitre, nous proposons un cadre de modélisation original dans lequel les volatilités de l'indice et de ses composantes sont reliées. Nous obtenons un modèle simplifié quand la taille de l'indice est grande, dans lequel l'indice suit un modèle à volatilité locale et les actions individuelles suivent un modèle à volatilité stochastique composé d'une partie intrinsèque et d'une partie commune dirigée par l'indice. Nous étudions la calibration de ces modèles et montrons qu'il est possible de se caler sur les prix d'options observés sur le marché, à la fois pour l'indice et pour les actions, ce qui constitue un avantage considérable. Enfin, dans le dernier chapitre de la thèse, nous développons un modèle à intensités permettant de modéliser simultanément, et de manière consistante, toutes les transitions de *ratings* qui surviennent dans un grand portefeuille de crédit. Afin de générer des niveaux de dépendance plus élevés, nous introduisons le modèle *dynamic frailty* dans lequel une variable dynamique inobservable agit de manière multiplicative sur les intensités de transitions. Notre approche est purement historique et nous étudions l'estimation par maximum de vraisemblance des paramètres de nos modèles sur la base de données de transitions de *ratings* passées.

Abstract

The first part of this thesis deals with probabilistic numerical methods for simulating the solution of a stochastic differential equation (SDE). We start with the algorithm of Beskos *et al.* [13] which allows exact simulation of the solution of a one dimensional SDE. We present an extension for the exact computation of expectations and we study the application of these techniques for the pricing of Asian options in the Black & Scholes model. Then, in the second chapter, we propose and study the convergence of two discretization schemes for a family of stochastic volatility models. The first one is well adapted for the pricing of *vanilla* options and the second one is efficient for the pricing of *path-dependent* options. We also study the particular case of an Orstein-Uhlenbeck process driving the volatility and we exhibit a third discretization scheme which has better convergence properties. Finally, in the third chapter, we tackle the trajectorial weak convergence of the Euler scheme by providing a simple proof for the estimation of the Wasserstein distance between the solution and its Euler scheme, uniformly in time.

The second part of the thesis is dedicated to the modelling of dependence in finance through two examples : the joint modelling of an index together with its composing stocks and intensity-based credit portfolio models. In the forth chapter, we propose a new modelling framework in which the volatility of an index and the volatilities of its composing stocks are connected. When the number of stocks is large, we obtain a simplified model consisting of a local volatility model for the index and a stochastic volatility model for the stocks composed of an intrinsic part and a systemic part driven by the index. We study the calibration of these models and show that it is possible to fit the market prices of both the index and the stocks. Finally, in the last chapter of the thesis, we define an intensity-based credit portfolio model. In order to obtain stronger dependence levels between rating transitions, we extend it by introducing an unobservable random process (*frailty*) which acts multiplicatively on the intensities of the firms of the portfolio. Our approach is fully historical and we estimate the parameters of our model to past rating transitions using maximum likelihood techniques.

Remerciements

Je tiens à remercier en premier lieu mon directeur de thèse, Benjamin Jourdain, pour tout le temps qu'il m'a accordé durant ces trois dernières années. Son encadrement exemplaire, sa rigueur scientifique, la qualité de ses relectures, sa constante bonne humeur ainsi que son soutien permanent ont été décisifs pour le bon déroulement de ma thèse. Je lui suis également très reconnaissant de m'avoir permis d'enseigner à l'École des Ponts. À ce titre, je voudrai aussi remercier Jean-François Delmas pour m'avoir permis d'intervenir dans le cours de probabilités de l'ENSTA.

Emmanuel Gobet et Antoine Lejay m'ont fait l'honneur d'accepter la rude tâche de rapporteur. Je les remercie pour leur lecture très attentive du manuscrit et leurs remarques toujours constructives. J'ai aussi été très honoré que Francesco Russo, Vlad Bally et Jean-David Fermanian aient accepté de faire partie de mon jury de thèse. Qu'ils trouvent ici l'expression de ma profonde gratitude.

Un grand merci à toute la famille du CERMICS, en particulier aux membres de l'équipe de Probabilités. Je commencerai par Aurélien, Bernard et Jean-François qui, chacun à sa façon, m'ont beaucoup aidé par leur conseils, encouragements et surtout par l'intérêt qu'ils ont porté à mes travaux. Merci à tous mes collègues doctorants pour tous les échanges scientifiques et humains que nous avons pu développer : je pense à Raphaël avec qui j'ai eu grand plaisir à partager le même bureau pendant les deux dernières années de ma thèse, à Abdelkoddous dont la bonne humeur contagieuse m'a souvent été bénéfique, à Jerome et Pierre pour nos innombrables discussions sur l'enseignement, l'informatique, la musique, le cinéma et bien d'autres sujets, mais aussi à tous ceux que j'ai côtoyés : Jean-Philippe, Julien, Simone, Piergiacomo, Cristina, Nadia, Infante, Maxence, Kimiya, Ronan, ...

Enfin, je tiens à exprimer ma plus profonde reconnaissance à ma famille et à mes amis pour leur soutien indéfectible et leur amour, avec une pensée particulière pour celle qui a toujours été mon moteur dans la vie : Emira.

Table des matières

Introduction **3**

I Méthodes de simulation exacte et schémas de discrétisation d'EDS. Applications en finance **23**

1 Méthodes de Monte Carlo exactes et application au pricing d'options asiatiques **25**
1.1 Exact Simulation techniques . . . 27
1.1.1 The exact simulation method of Beskos *et al.* [13] . . . 27
1.1.2 The unbiased estimator (U.E) . . . 30
1.2 Application : the pricing of continuous Asian options . . . 32
1.2.1 The case $\alpha \neq 0$ . . . 34
1.2.2 Standard Asian options : the case $\alpha = 0$ and $\beta > 0$ . . . 38
1.3 Conclusion . . . 48
1.4 Appendix . . . 49
1.4.1 The practical choice of p and q in the U.E method . . . 49
1.4.2 Simulation from the distribution h given by (1.13) . . . 50

2 Schémas de discrétisation pour modèles à volatilité stochastique **53**
2.1 An efficient scheme for path dependent options pricing . . . 56
2.1.1 General case . . . 57
2.1.2 Special case of an Ornstein-Uhlenbeck process driving the volatility . . . 64
2.2 A second order weak scheme . . . 71
2.3 Numerical results . . . 77
2.3.1 Numerical illustration of strong convergence properties . . . 78
2.3.2 Standard call pricing . . . 81
2.3.3 Asian option pricing and multilevel Monte Carlo . . . 82
2.4 Conclusion . . . 84
2.5 Appendix . . . 84
2.5.1 Proof of Lemma 21 . . . 84
2.5.2 Proof of Lemma 26 . . . 85
2.5.3 Proof of Lemma 32 . . . 85
2.5.4 Proof of Lemma 33 . . . 86

3 Erreur faible uniforme en temps pour le schéma d'Euler **87**
3.1 Introduction . . . 87
3.2 Résultat principal . . . 90
3.3 Résultats auxiliaires . . . 91
3.4 Preuve du Théorème 7 . . . 94
3.5 Preuve de la Proposition 12 . . . 94
3.5.1 Estimation de $\left|\int_0^t \Delta_1(s)ds\right|$ . . . 96
3.5.2 Estimation de $\left|\int_0^t \Delta_2(s)ds\right|$ . . . 99

II Modélisation de la dépendance en finance : modèle d'indices boursiers et modèles de portefeuilles de crédit **105**

4 Un modèle couplant indice et actions **107**
4.1 Model Specification . . . 108
4.2 Asymptotics for a large number of underlying stocks . . . 110
4.3 Model calibration . . . 114
4.3.1 Simplified model . . . 114
4.3.2 Original model . . . 124
4.4 Illustration of Theorems 35 and 36 and comparison with a constant correlation model 124
4.4.1 Application: Pricing of a worst-of option . . . 128
4.5 Conclusion . . . 129
4.6 Appendix . . . 130
4.6.1 Proof of Theorem 35 . . . 131
4.6.2 Proof of Theorem 36 . . . 133

5 Estimation d'un modèle à intensités pour la gestion des risques. Extension aux modèles de *frailty* dynamique **135**
5.1 Introduction . . . 136
5.2 The basic model . . . 137
5.3 Computation of the transition matrices and Tests in sample . . . 144
5.4 Extension to frailty models . . . 144
5.5 Conclusion . . . 149

Bibliographie **156**

Introduction

La thèse que je présente se décompose en deux parties indépendantes mais qui s'inscrivent toutes les deux dans le cadre des mathématiques appliquées à la finance. La première partie est consacrée à l'étude mathématique de méthodes numériques pour la simulation de processus aléatoires définis par des équations différentielles stochastiques (notées EDS ci-après), et à leurs applications en finance. La deuxième partie porte plutôt sur des aspects de modélisation. Plus précisément, nous nous sommes intéressés à la modélisation de la dépendance en finance à travers deux problématiques distinctes : la modélisation jointe entre un indice boursier et les actifs qui le composent et la gestion du risque pour les portefeuilles de crédit.

Ce premier chapitre introductif a pour objectif de présenter les enjeux et les principaux résultats de la thèse, en évitant, autant que faire se peut, d'entrer dans les détails techniques qui, eux, seront développés par la suite.

1. Simulation d'EDS et applications en finance

En tant qu'objet mathématique, les équations différentielles stochastiques doivent leur essor au mathématicien japonais Kiyoshi Itô qui a posé les jalons théoriques de l'intégrale stochastique et des règles de calcul y afférant. Leur utilisation en tant qu'outil mathématique pour la modélisation en finance s'est largement répandue ces dernières décennies, notamment depuis le fameux modèle de Black & Scholes. Dans ce dernier, sous la probabilité risque neutre, unique en l'occurrence, le prix $(S_t)_{t\geq 0}$ d'une action cotée en bourse suit une EDS linéaire à coefficients constants :

$$dS_t = rS_t dt + \sigma S_t dW_t,$$

σ et r représentant respectivement la volatilité et le taux d'intérêt sans risque et $(W_t)_{t\geq 0}$ désignant un mouvement Brownien réel.

Outre la complétude, un des avantages de ce modèle, et ce qui explique en bonne partie le succès qu'il a rencontré, est le fait que l'on dispose d'une solution explicite pour le prix sous la probabilité risque neutre, à savoir $S_t = S_0 e^{\sigma W_t + (r - \frac{\sigma^2}{2})t}$, permettant de mener à bien plusieurs calculs importants en pratique : calcul des prix d'options européennes (*Call*, *Put*, digitales,...), des sensibilités de ces prix par rapport aux paramètres (les grecques), des prix de certaines options exotiques (options barrières, option *lookback*, ...), etc. Toutefois, le modèle de Black & Scholes n'est pas exempt de critiques et il est avéré depuis longtemps que les hypothèses sous-jacentes à ce dernier ne sont pas en adéquation avec les marchés financiers, surtout la constance de la volatilité. D'où l'émergence de nouveaux modèles, beaucoup plus réalistes, comme les modèles à volatilité stochastiques, où la volatilité est supposée suivre une EDS autonome éventuellement corrélée avec celle qui gouverne le cours de l'action, ou encore les modèles à volatilité locale, où la volatilité est fonction du temps et du cours de l'action [1]. Malheureusement, il est alors rare de tomber sur des EDS qui admettent des solutions explicites, ce qui justifie le besoin de recourir aux méthodes numériques.

Plus généralement, il arrive souvent, en finance comme en d'autres domaines d'application des mathématiques, que l'on cherche à calculer des quantités qui s'écrivent sous la forme

$$\mathbb{E}\left[f\left((X_t)_{t\in[0,T]}\right)\right], \tag{1}$$

où f est une fonctionnelle donnée et le processus $(X_t)_{t\in[0,T]}$ est la solution d'une EDS que l'on ne sait pas résoudre explicitement. Pour un probabiliste, qui dit espérance dit méthodes de Monte

1. Nous pouvons aussi citer les modèles à sauts mais cela ne rentre pas dans le cadre de cette thèse dans la mesure où les méthodes numériques que j'ai étudiées ne peuvent s'appliquer qu'aux modèles continus.

Carlo. J'ai donc consacré une bonne partie de ma thèse à la proposition et à l'étude de méthodes numériques probabilistes permettant de répondre à ce type de problématique.

Ainsi, le premier chapitre s'articule autour de la méthode de simulation exacte de Beskos *et al.* [13], de son extension à des fins de calcul exact d'espérances et de l'application de ces idées à l'évaluation du prix d'options asiatiques dans le modèle de Black & Scholes. Le mot "exact" ici fait opposition aux schémas de discrétisation des EDS qui, en plus de l'erreur statistique provenant de l'approximation de l'espérance par une méthode de Monte Carlo, introduisent justement un biais de discrétisation. Le deuxième chapitre s'attache à la proposition et à l'étude de la convergence de nouveaux schémas de discrétisation pour une famille de modèles à volatilité stochastique. Enfin, dans le troisième chapitre, nous apportons une première réponse à l'étude de la convergence faible trajectorielle d'un schéma de discrétisation, fameux s'il en est : le schéma d'Euler.

1.1 Méthodes de Monte Carlo exactes. Application au pricing d'options asiatiques

Ce premier chapitre correspond à un article écrit avec mon directeur de thèse Benjamin Jourdain (cf. Jourdain et Sbai [60]). Il a été publié dans la revue *Monte Carlo Methods and Applications*.

1.1.1 Méthodes de Monte Carlo exactes

Récemment, Beskos *et al.* [13] ont introduit un algorithme original permettant de simuler exactement les trajectoires d'un processus solution d'une EDS en dimension 1. L'idée de base consiste à simuler un tel processus par une méthode d'acceptation-rejet qui utilise comme loi de proposition la loi du mouvement Brownien. Pour ce faire, plusieurs étapes intermédiaires sont nécessaires. La première partie du chapitre 1 s'attache à décrire la méthodologie de Beskos *et al.* [13] dans un cadre mathématique rigoureux. Sans rentrer dans les détails, rappelons rapidement le fonctionnement de cet algorithme :

- Quitte à faire un changement de variable (une transformation de Lamperti), on part de l'EDS unidimensionnelle suivante :

$$\left\{ \begin{array}{rcl} dX_t & = & a(X_t)dt + dW_t \\ X_0 & = & x. \end{array} \right. \tag{2}$$

- Sous certaines hypothèses, on peut trouver un processus $(Z_t)_{t\in[0,T]}$ qui, conditionnellement à sa valeur terminale, possède la même loi que $(W_t^x)_{t\in[0,T]}$, le mouvement Brownien issu de x, et une fonction ϕ positive qui dépend du drift a, tels que la loi de $(X_t)_{t\in[0,T]}$ soit absolument continue par rapport à celle de $(Z_t)_{t\in[0,T]}$ et que sa dérivée de Radon-Nikodym soit proportionnelle à $\exp\big[-\int_0^T \phi(Z_t)\,dt\big]$.
- On simule un événement de probabilité $\exp\big[-\int_0^T \phi(Z_t)\,dt\big]$ à l'aide d'un processus de Poisson ponctuel :
  - Soit $(Z_t(\omega))_{t\in[0,T]}$ une réalisation du processus $(Z_t)_{t\in[0,T]}$ et soit $M(\omega)$ une borne supérieure de la fonction $t \in [0,T] \mapsto \phi(Z_t(\omega))$.
  - Soient $N \sim \mathcal{P}\big(TM(\omega)\big)$ une variable aléatoire qui suit une loi de Poisson de paramètre $TM(\omega)$ et, indépendamment, $(U_i, V_i)_{i=1\dots N} \overset{\text{i.i.d}}{\sim} \mathcal{U}\big([0,T]\times[0,M(\omega)]\big)$ une suite de points aléatoires indépendants uniformément répartis dans le rectangle $[0,T]\times[0,M(\omega)]$.

On a alors

$$\mathbb{P}\left(\#\{i \leq N, V_i \leq \phi(Z_{U_i}(\omega))\} = 0\right) = \exp\left[-\int_0^T \phi(Z_t(\omega))\, dt\right].$$

Il suffit donc de simuler le processus $(Z_t)_{t\in[0,T]}$ aux instants $(U_i)_{i=1\ldots N}$. La trajectoire est acceptée si, pour tout $1 \leq i \leq N$, $V_i \geq \phi(Z_{U_i}(\omega))$. Elle est rejetée sinon (voir Figure 1).

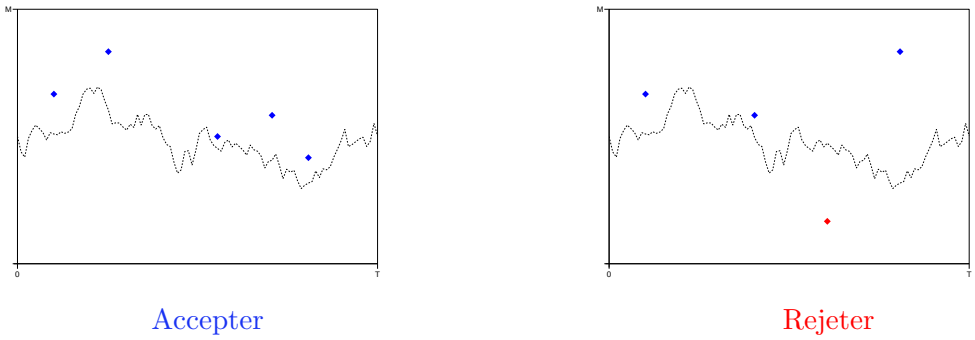

FIGURE 1 – Illustration de l'algorithme de Beskos *et al.* [13]

Afin de mettre en oeuvre cet algorithme, il est nécessaire de pouvoir spécifier le rectangle fini au sein duquel on simule le processus de Poisson ponctuel, c'est à dire la borne $M(w)$. Dans un premier article, Beskos et Roberts [15] supposent que la fonction ϕ est majorée. Beskos *et al.* [13] assouplissent cette dernière hypothèse en supposant que

$$\limsup_{u\to+\infty} \phi(u) < +\infty \text{ ou } \limsup_{u\to-\infty} \phi(u) < +\infty.$$

La simulation se fait alors en simulant le mouvement Brownien de manière récursive, conditionnellement à sa valeur terminale et à son minimum ou maximum.

Toutefois, cette dernière hypothèse reste assez restrictive en pratique. De plus, en finance ce n'est pas tellement la simulation des processus qui est importante mais le calcul d'espérance. Dans cette optique, nous proposons une méthode de calcul exact d'espérance qui s'appuie sur l'algorithme qu'on vient de décrire. L'idée est d'utiliser le développement en série de l'exponentielle et de faire apparaître l'espérance d'une variable aléatoire discrète.

Plus précisément, on cherche à calculer

$$C_0 = \mathbb{E}\left(f(X_T)\right),$$

où $(X_t)_{t\in[0,T]}$ est solution de l'EDS (2). En s'inspirant de l'algorithme de Beskos *et al.* [13], on montre qu'il existe deux fonctions ψ et ϕ et un processus $(Z_t)_{t\in[0,T]}$ que l'on sait simuler tels que

$$C_0 = \mathbb{E}\left(\psi(Z_T)\exp\left[-\int_0^T \phi(Z_t)dt\right]\right). \tag{3}$$

Sous une condition d'intégrabilité renforcée, nous construisons alors un estimateur sans biais de C_0, facilement simulable, qui s'écrit sous la forme :

$$\psi(Z_T)e^{-c_Z T}\frac{1}{p_Z(N)N!}\prod_{i=1}^{N}\frac{c_Z - \phi(Z_{V_i})}{q_Z(V_i)}, \tag{4}$$

où c_Z est une variable aléatoire mesurable par rapport à la tribu engendrée par le processus $(Z_t)_{t\in[0,T]}$ et, conditionnellement à ce dernier,

- N est une variable aléatoire discrète de loi p_Z strictement positive.
- $(V_i)_{i\in\mathbb{N}^*}$ est une suite de variables aléatoires à valeurs dans $[0,T]$, indépendantes et identiquement distribuées suivant une loi q_Z strictement positive.
- N et $(V_i)_{i\in\mathbb{N}^*}$ sont indépendantes.

L'idée derrière ce type d'estimation remonte à Wagner [129]. Plus récemment, Beskos *et al.* [14] et Fearnhead *et al.* [38]) ont introduit deux versions particulières de cet estimateur : le *Poisson estimator* et le *Generalized Poisson estimator*.

Nous montrons que cette méthode de calcul exact d'espérances est une extension de la méthode de simulation exacte de Beskos *et al.* [13] et nous explorons les possibles méthodes de réduction de variance que l'on peut appliquer.

1.1.2 Application au pricing d'options asiatiques

Dans la deuxième partie du chapitre, nous adaptons les méthodes exactes citées précédemment pour le calcul du prix de certaines options exotiques dans le cadre du modèle de Black & Scholes. L'intérêt par rapport à une méthode de Monte Carlo classique est que l'on évite le biais de discrétisation résultant de l'utilisation de schémas numériques pour les EDS.

Ainsi, nous montrons comment appliquer la méthode de simulation exacte et la méthode de calcul exact d'espérance pour calculer le prix d'une option de pay-off $f\left(\alpha S_T + \beta \int_0^T S_t dt\right)$ avec α et β deux constantes positives et $(S_t)_{t\in[0,T]}$ le prix d'une action dans le modèle de Black & Scholes. Pour $\alpha > 0$, nous montrons que $\alpha S_T + \beta \int_0^T S_t dt$ a même loi que la solution à l'instant T d'une EDS unidimensionnelle bien choisie, grâce à un changement de variables inspiré de Rogers et Shi [105]. Les résultats numériques obtenus montrent entre autres que notre méthode de calcul exact d'espérances est plus compétitive qu'une méthode de Monte Carlo classique basée sur un schéma de discrétisation.

Plus particulièrement, nous considérons une option asiatique classique, ce qui correspond au cas $\alpha = 0$. En l'occurrence, le changement de variables précédent n'est plus valide et nous proposons un nouveau changement de variables qui présente une singularité à l'instant initial. Par conséquent, sa loi n'est pas absolument continue par rapport à celle du mouvement Brownien et nous introduisons un nouveau processus gaussien. Nous montrons que les conditions d'application des méthodes exactes ne sont pas réunies et nous proposons une méthode hybride qui marie simulation exacte par rejet et développement en série entière de l'exponentielle pour calculer le prix de l'option asiatique.

1.2 Schémas de discrétisation pour modèles à volatilité stochastique

1.2.1 Motivations

La simulation exacte des solutions d'EDS n'est pas toujours possible. Par exemple, en dimension supérieure à un, il n'est pas évident de pouvoir se ramener à un coefficient de diffusion constant. De plus, l'approche que l'on vient de décrire n'est possible que lorsque le coefficient de dérive s'écrit comme le gradient d'une fonction : ce qui est gratuit en dimension un devient très restrictif en grande dimension. En fait, afin de calculer des espérances de fonctionnelles de la solution d'une EDS avec une méthode de Monte Carlo, on se tourne habituellement vers les schémas de discrétisation.

En finance, les modèles à volatilité stochastiques sont un exemple pertinent de l'utilité de ces méthodes numériques. En effet, les EDS en dimension deux qui définissent ces modèles sont rarement simulables de manière exacte. Exception faite du modèle de Heston [55] pour lequel Broadie et Kaya [20] proposent une méthode de simulation exacte mais qui s'avère coûteuse en temps. Dans le chapitre 2, on se propose de construire et d'analyser la vitesse de convergence de schémas de discrétisation efficaces pour une famille de modèles à volatilité stochastique.

Avant de préciser les résultats de cette partie de la thèse, commençons par présenter quelques résultats connus sur la discrétisation des EDS, un sujet qui a servi, et qui sert toujours, de matière à une vaste littérature.

Quand on veut utiliser un schéma $(\widetilde{X}_t^N)_{t\in[0,T]}$ pour calculer par une méthode de Monte Carlo des quantités du type $\mathbb{E}\left(f(X_T)\right)$, où $(X_t)_{t\in[0,T]}$ est la solution d'une EDS, le critère de convergence qu'il faut regarder est l'erreur faible, c'est à dire l'erreur en loi à l'instant terminal. Plus précisément, on s'intéresse au comportement en fonction du pas de temps de $\left|\mathbb{E}\left(f(X_T)\right) - \mathbb{E}(f(\widetilde{X}_T^N))\right|$ pour une classe assez large de fonctions tests f. Cette problématique se rencontre souvent en finance, notamment lorsqu'il s'agit d'évaluer le prix ou de couvrir des options *vanilles*.

Le schéma numérique le plus couramment utilisé et le plus largement étudié est sans doute le schéma d'Euler. Talay et Tubaro [117] ont montré que l'erreur faible de ce schéma, pour des fonctions à croissance polynômiale, admet un développement limité en fonction du pas de discrétisation, ce qui permet d'appliquer la méthode d'extrapolation de Romberg pour accélérer la convergence. Bally et Talay [7] et Guyon [52] ont généralisé ce résultat à une classe plus large de fonctions tests, respectivement les fonctions mesurables bornées et les distributions tempérées. Le terme principal du développement de l'erreur est en $\frac{1}{N}$ où N est le nombre de pas de discrétisation. On dit alors que le schéma d'Euler est d'ordre faible 1. On trouve également dans la littérature des schémas d'ordre faible plus élevé. Par exemple, Kusuoka [76, 77] introduit des schémas d'ordre faible arbitrairement élevé en remplaçant les intégrales itérées qui apparaissent dans les développements de Taylor stochastiques par des variables aléatoires définies sur un espace fini et qui préservent les moments jusqu'à un certain ordre (voir aussi Ninomiya [95, 96] pour l'implémentation de ces schémas et leur application en finance). Citons également les schémas d'ordre faible deux de Ninomiya et Victoir [98] et de Ninomiya et Ninomiya [97] qui utilisent des équations différentielles ordinaires ou encore les formules de cubature sur l'espace de Wiener obtenues par Lyons et Victoir [87].

Par ailleurs, l'analyse mathématique de la convergence des schémas de discrétisation ne se limite pas à l'erreur faible. On étudie aussi l'erreur forte, c'est à dire la distance, pour une même source d'aléa, entre la trajectoire de la solution et celle de son schéma. Généralement, on regarde la norme L^2 sur l'espace des trajectoires : $\sqrt{\mathbb{E}\left[\sup_{t\in[0,T]}\left\|X_t - \widetilde{X}_t^N\right\|^2\right]}$.

Il est bien connu que l'erreur forte du schéma d'Euler est en $\frac{1}{\sqrt{N}}$, c'est à dire d'ordre $\frac{1}{2}$. Le schéma de Milstein est un schéma d'ordre 1. Toutefois, en dimension supérieure à un, il fait intervenir des intégrales stochastiques du type $\int_0^t W_s\, dB_s$ pour $(W_t)_{t\in[0,T]}$ et $(B_t)_{t\in[0,T]}$ deux mouvements Browniens indépendants, ce qu'on ne sait pas simuler de manière exacte. Il y a moyen d'éviter cela si une condition de commutativité restrictive est satisfaite.

Il arrive aussi que l'on cherche à calculer des espérances de fonctionnelles de la trajectoire, auquel cas il est plus judicieux de regarder l'erreur faible trajectorielle, c'est à dire l'erreur en loi sur toute la trajectoire. On se pose alors la question suivante : pour une large classe de fonctionnelles f, quel est le comportement de $\left|\mathbb{E}\left(f((X_t)_{t\in[0,T]})\right) - \mathbb{E}(f((\widetilde{X}_t^N)_{t\in[0,T]}))\right|$ en fonction du pas de discrétisation $\frac{T}{N}$?

C'est le cas par exemple en finance avec les options dites *path-dependent*, c'est-à-dire celles dont le prix dépend de toute la trajectoire de l'actif sous-jacent et non seulement de sa valeur terminale.

Par une approche très originale, Cruzeiro *et al.* [27] obtiennent un schéma de discrétisation dont l'erreur faible trajectorielle est d'ordre un : sous hypothèse d'ellipticité, il est possible de trouver une rotation intelligente du mouvement Brownien qui gouverne l'EDS de telle sorte que le schéma de Milstein ne fait plus intervenir des intégrales itérées et devient facilement simulable.

Enfin, il est utile de noter que, malgré la particularité des EDS qui régissent les modèles à volatilité stochastique, il existe relativement peu de travaux sur des schémas de discrétisation spécifiquement adaptés à ces EDS. Exceptionnellement, le modèle de Heston [55], en particulier le processus CIR qui dirige la volatilité, a reçu une attention particulière : voir par exemple Deelstra et Delbaen [29], Alfonsi [1], Kahl et Schurz [62], Andersen [3], Berkaoui *et al.* [12], Ninomiya et Victoir [98], Lord *et al.* [86] et Alfonsi [2]. Mentionnons aussi l'article de Kahl et Jäckel [61] qui étudient différents schémas numériques pour modèles à volatilité stochastique et qui obtiennent un schéma d'ordre fort $\frac{1}{2}$ mais avec une constante multiplicative meilleure que celle du schéma d'Euler.

1.2.2 Résultats

Nous considérons le modèle de volatilité stochastique suivant pour un actif $(S_t)_{t\in[0,T]}$

$$\left\{\begin{array}{lcll} dS_t & = & rS_t dt + f(Y_t)S_t\left(\rho dW_t + \sqrt{1-\rho^2}dB_t\right); & S_0 = s_0 > 0 \\ dY_t & = & b(Y_t)dt + \sigma(Y_t)dW_t; & Y_0 = y_0 \end{array}\right. , \tag{5}$$

où r représente le taux d'intérêt sans risque, $(B_t)_{t\in[0,T]}$ et $(W_t)_{t\in[0,T]}$ sont deux mouvements Browniens indépendants, $\rho \in [-1,1]$ est un coefficient de corrélation constant et f est une fonction positive strictement monotone.

Cette spécification englobe plusieurs modèles à volatilité stochastique connus : les modèles de Hull et White [57], de Scott [112], de Stein et Stein [115], de Heston [55] ou encore les modèles quadratiques gaussiens. Nous supposerons que les fonctions f et σ sont régulières, plus précisément nous travaillerons sous l'hypothèse suivante tout le long du chapitre :

$$(\mathcal{H}) \quad f \text{ et } \sigma \text{ sont des fonctions } \mathcal{C}^1 \text{ et } \sigma > 0.$$

Nous ne traitons donc pas le modèle de Heston.

Nous allons tirer profit de la structure particulière de l'EDS bidimensionnelle (5) : le processus $(Y_t)_{t\in[0,T]}$ qui dirige la volatilité suit une EDS autonome donc, en utilisant la même astuce qui a servi à se débarrasser de l'intégrale stochastique dans la méthode de simulation exacte précédemment décrite, on se débarrasse de l'intégrale stochastique par rapport au mouvement Brownien commun $(W_t)_{t\in[0,T]}$ dans l'EDS qui dirige l'actif.

La mise en oeuvre de cette approche se traduit par l'obtention de l'équation suivante pour le couple $(X_t, Y_t)_{t\in[0,T]}$ où $X_t = \log(S_t)$:

$$\left\{\begin{array}{lcl} dX_t & = & \rho dF(Y_t) + h(Y_t)dt + \sqrt{1-\rho^2}f(Y_t)dB_t \\ dY_t & = & b(Y_t)dt + \sigma(Y_t)dW_t \end{array}\right. , \tag{6}$$

avec $F : y \mapsto \int_0^y \frac{f}{\sigma}(z)dz$ et $h : y \mapsto r - \frac{1}{2}f^2(y) - \rho(\frac{b}{\sigma}f + \frac{1}{2}(\sigma f' - f\sigma'))(y)$.

En se basant sur cette transformation, nous proposons deux schémas de discrétisation et en étudions la convergence. Le premier, basé sur le schéma de Milstein pour le processus $(Y_t)_{t\in[0,T]}$,

possède une erreur de convergence faible trajectorielle d'ordre un. Le deuxième, basé sur le schéma de Ninomiya et Victoir [98] pour le processus $(Y_t)_{t\in[0,T]}$, possède une erreur de convergence faible d'ordre deux. Le cas particulier où $(Y_t)_{t\in[0,T]}$ est un processus d'Ornstein-Uhlenbeck fait l'objet d'un traitement spécifique et nous exhibons un schéma de discrétisation qui possède de bonnes propriétés de convergence, à la fois pour la convergence faible et pour la convergence faible trajectorielle. Précisons tout cela.

Sur l'intervalle de temps $[0,T]$, on considère la grille de discrétisation uniforme de pas $\delta_N = \frac{T}{N}$ pour $N \in \mathbb{N}^* : t_k = k\delta_N, 0 \leq k \leq N$. Pour simplifier les notations, introduisons la fonction $\psi : y \mapsto f^2(y)$, $\underline{\psi}$ sa borne inférieure et $\overline{\psi}$ sa borne supérieure.

1. <u>Un schéma pour les options *path-dependent*</u>

   On introduit le schéma suivant : $\widetilde{X}_0^N = \log(s_0)$ et $\forall 0 \leq k \leq N-1$,

$$\begin{aligned}\widetilde{X}_{t_{k+1}}^N &= \widetilde{X}_{t_k}^N + \rho\left(F(\widetilde{Y}_{t_{k+1}}^N) - F(\widetilde{Y}_{t_k}^N)\right) + \delta_N h(\widetilde{Y}_{t_k}^N) \\ &+\sqrt{1-\rho^2}\sqrt{\left(\psi(\widetilde{Y}_{t_k}^N) + \frac{\sigma\psi'(\widetilde{Y}_{t_k}^N)}{\delta_N}\int_{t_k}^{t_{k+1}}(W_s - W_{t_k})ds\right) \vee \underline{\psi}}\,\Delta B_{k+1}\end{aligned} \tag{7}$$

   où on note par $\Delta B_{k+1} = B_{t_{k+1}} - B_{t_k}$ l'accroissement du mouvement Brownien $(B_t)_{t\in[0,T]}$ et par $(\widetilde{Y}_t^N)_{t\in[0,T]}$ le schéma de Milstein de $(Y_t)_{t\in[0,T]}$. On montre que la convergence faible trajectorielle de ce schéma est d'ordre un. Plus précisément, on montre le résultat suivant :

   Théorème 1 *Supposons que*
   – *b et σ sont respectivement $\mathcal{C}^3$ et $\mathcal{C}^4$, bornées avec des dérivées bornées et avec $\inf_{y\in\mathbb{R}} \sigma(y) > 0$.*
   – *f est $\mathcal{C}^4$, bornée avec des dérivées bornées.*
   – *$\underline{\psi} > 0$.*
   Alors, pour tout $p \geq 1$, il existe une constante $C_p > 0$ indépendante de N tel que

$$\mathbb{E}\left[\max_{0\leq k\leq N}\left\|\left(\widetilde{X}_{t_k}, Y_{t_k}\right) - \left(\widetilde{X}_{t_k}^N, \widetilde{Y}_{t_k}^N\right)\right\|^{2p}\right] \leq \frac{C_p}{N^{2p}}.$$

   où $(\widetilde{X}_{t_0}, \ldots, \widetilde{X}_{t_N})$ est un vecteur aléatoire qui a même loi que $(X_{t_0}, \ldots, X_{t_N})$, défini par $\widetilde{X}_{t_0} = X_{t_0}$ et, $\forall 0 \leq k < N$,

$$\widetilde{X}_{t_{k+1}} = \widetilde{X}_{t_k} + \rho(F(Y_{t_{k+1}}) - F(Y_{t_k})) + \int_{t_k}^{t_{k+1}} h(Y_s)ds + \sqrt{\frac{1-\rho^2}{\delta_N}\int_{t_k}^{t_{k+1}}\psi(Y_s)ds}\,\Delta B_{k+1}.$$

   On s'intéresse aussi au cas particulier où le processus qui dirige la volatilité est un processus d'Ornstein-Uhlenbeck, c'est-à-dire quand $(Y_t)_{t\in[0,T]}$ est solution de l'EDS suivante :

$$dY_t = \nu dW_t + \kappa(\theta - Y_t)dt \tag{8}$$

   Il est alors possible de simuler exactement ce processus et on montre que si on remplace le schéma de Milstein par la solution exacte dans le schéma (7), on préserve l'ordre de convergence. On réussit même à assouplir les hypothèses du théorème (1), en particulier l'hypothèse

$\underline{\psi} > 0$, ce qui permet de traiter le modèle de Scott [112] et donc celui de Hull et White [57] également.

Le fait de pouvoir profiter de la simulation exacte de $(Y_t)_{t\in[0,T]}$ sans altérer l'ordre de convergence est un avantage de notre schéma par rapport au schéma de Cruzeiro *et al.* [27]. Mieux encore, on montre que notre schéma est plus adapté à la méthode *multilevel Monte Carlo* introduite par Giles [48]. Précisons rapidement notre propos.

La méthode *multilevel Monte Carlo*, qui est une généralisation de la méthode de Romberg statistique de Kebaier [65], permet de calculer de manière efficace l'espérance d'une fonctionnelle de la solution d'une EDS par une méthode Monte Carlo. L'idée consiste à combiner les estimations basées sur un même schéma de discrétisation avec des pas de discrétisation différents de manière à réduire la complexité permettant d'atteindre une précision donnée. L'efficacité de cette méthode repose essentiellement sur la vitesse de convergence forte du schéma en question, plus précisément sur l'erreur forte entre le schéma de pas grossier et le schéma de pas plus fin. Par exemple, pour calculer l'espérance d'une fonctionnelle lipschitzienne de la trajectoire en utilisant un schéma d'ordre fort 1, la méthode *multilevel Monte Carlo* permet de réduire le coût de calcul pour atteindre une précision $\epsilon > 0$ de $\mathcal{O}(\epsilon^{-3})$ à $\mathcal{O}(\epsilon^{-2})$. Nous montrons comment coupler notre schéma de pas $\frac{T}{N}$ avec celui de pas $\frac{T}{2N}$ de manière à avoir une erreur forte d'ordre 1. C'est la structure particulière de notre schéma qui rend un tel couplage possible, ce que ne permet pas de faire le schéma de Cruzeiro *et al.* [27].

2. <u>Un schéma pour les options *vanilles*</u>

En remarquant que, conditionnellement à $(Y_t)_{t\in[0,T]}$,

$$X_T \sim \mathcal{N}\left(\log(s_0) + \rho(F(Y_T) - F(y_0)) + \int_0^T h(Y_s)ds\,,\,(1-\rho^2)\int_0^T f^2(Y_s)ds\right)$$

on propose le schéma de discrétisation suivant

$$\begin{aligned}\overline{X}_T^N = \log(s_0) + \rho(F(\overline{Y}_T^N) - F(y_0)) + \delta_N \sum_{k=0}^{N-1} \frac{h(\overline{Y}_{t_k}^N) + h(\overline{Y}_{t_{k+1}}^N)}{2} \\ + \sqrt{(1-\rho^2)\delta_N \sum_{k=0}^{N-1} \frac{f^2(\overline{Y}_{t_k}^N) + f^2(\overline{Y}_{t_{k+1}}^N)}{2}}\; G\end{aligned} \tag{9}$$

où $(\overline{Y}_{t_k}^N)_{0\le k\le N}$ est le schéma de Ninomiya-Victoir de $(Y_t)_{t\in[0,T]}$ et G est une gaussienne indépendante centrée réduite.

On montre alors le résultat suivant :

Théorème 2 *Si on a*

- $|\rho| \neq 1$,
- *f et h des fonctions $\mathcal{C}^4$ bornées et avec des dérivées bornées. F une fonction $\mathcal{C}^6$ bornée avec des dérivées bornées,*
- *b et σ respectivement $\mathcal{C}^4$ et $\mathcal{C}^5$ avec des dérivées bornées,*
- $\underline{\psi} > 0$,

alors, pour toute fonction g vérifiant $\exists c \ge 0, \mu \in [0,2)$ tel que $\forall y > 0, |g(y)| \le ce^{|\log(y)|^\mu}$, il existe $C > 0$ tel que

$$\left|\mathbb{E}\left(g\left(S_T\right)\right) - \mathbb{E}\left(g\left(e^{\overline{X}_T^N}\right)\right)\right| \le \frac{C}{N^2}.$$

3. Un schéma performant dans le cas Ornstein-Uhlenbeck
En s'inspirant du schéma d'ordre fort $\frac{3}{2}$ de Lapeyre et Temam [81] pour l'évaluation du prix des options asiatiques, nous proposons le schéma suivant dans le cas particulier où $(Y_t)_{t\in[0,T]}$ est solution de l'EDS (8) :

$$\widehat{X}^N_{t_{k+1}} = \widehat{X}^N_{t_k} + \rho\left(F(Y_{t_{k+1}}) - F(Y_{t_k})\right) + \widehat{h}_k + \sqrt{1-\rho^2}\sqrt{\widehat{\psi}_k}\ \Delta B_{k+1}, \tag{10}$$

avec $\widehat{h}_k = \delta_N h(Y_{t_k}) + \nu h'(Y_{t_k}) \int_{t_k}^{t_{k+1}} (W_s - W_{t_k})ds + (\kappa(\theta - Y_{t_k})h'(Y_{t_k}) + \frac{\nu^2}{2}h''(Y_{t_k}))\frac{\delta_N^2}{2}$ et $\widehat{\psi}_k = \left(\psi(Y_{t_k}) + \frac{\nu\psi'(Y_{t_k})}{\delta_N}\int_{t_k}^{t_{k+1}}(W_s - W_{t_k})ds + (\kappa(\theta - Y_{t_k})\psi'(Y_{t_k}) + \frac{\nu^2}{2}\psi''(Y_{t_k}))\frac{\delta_N}{2}\right) \vee \underline{\psi}$.
On vérifie alors que ce schéma a de bonnes propriétés de convergence, tant pour les options *vanilles* que pour les options *path-dependent*. Plus précisément, il possède un ordre de convergence faible égal à deux pour l'actif et un ordre de convergence faible trajectorielle égal à $\frac{3}{2}$ pour le triplet $\left(Y_t, \int_0^t h(Y_s)ds, \int_0^t f^2(Y_s)ds\right)_{t\in[0,T]}$ ce qui permet une amélioration considérable de la méthode *multilevel Monte Carlo*.

Dans la dernière partie de ce chapitre, nous effectuons plusieurs simulations numériques qui viennent corroborer les résultats théoriques obtenus. Nous illustrons aussi le gain réalisé, en termes de temps de calcul, quand on utilise nos différents schémas avec la méthode de *multilevel Monte Carlo* et ce à travers deux exemples pratiques : le pricing d'un call standard et d'une option asiatique dans le modèle de Scott. Comparées à ceux du schéma d'Euler, de Kahl et Jäckel [61] et de Cruzeiro *et al.* [27], les performances de nos schémas sont globalement très satisfaisantes.

1.3 Convergence faible uniforme en temps pour le schéma d'Euler

Soit l'EDS d-dimensionnelle suivante, $d \geq 1$:

$$\begin{cases} dX_t = b(X_t)dt + \sigma(X_t)dW_t \\ X_0 = x \in \mathbb{R}^d \end{cases}, \tag{11}$$

où $(W_t)_{t\in[0,T]}$ est un mouvement Brownien de dimension $r \geq 1$, $b : \mathbb{R}^d \to \mathbb{R}^d$ et $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times r}$. On désigne par $(X_t^x)_{t\in[0,T]}$ la solution de (11) partant de x et par $(X_t^{x,n})_{t\in[0,T]}$ son schéma d'Euler, n étant le nombre de points de discrétisation de l'intervalle $[0,T]$.

Le troisième chapitre de la thèse est consacré à l'étude de la convergence du schéma d'Euler. Comme il a été indiqué, ce schéma a fait l'objet d'une recherche abondante. Nous avons aujourd'hui une connaissance de plus en plus approfondie de la convergence faible de ce schéma mais nous connaissons relativement peu de résultats sur la convergence faible trajectorielle. Typiquement, la question suivante reste ouverte : pour une fonctionnelle $f : \mathcal{C}([0,T]) \to \mathbb{R}$ quelconque, quelle est le comportement de $\left|\mathbb{E}\left(f\left((X_t^x)_{t\in[0,T]}\right) - f\left((X_t^{x,n})_{t\in[0,T]}\right)\right)\right|$ en fonction du pas de discrétisation $\frac{T}{n}$?

On peut trouver dans la littérature des travaux qui abordent cette question pour des fonctionnelles particulières, généralement motivés par des exemples provenant de la finance de marché. Par exemple, Gobet [49] traite le cas des options barrières en montrant que cette vitesse est en $\frac{1}{n}$ pour les fonctionnelles du type $\mathbb{1}_{\{\forall 0\leq t\leq T, X_t^x\in\mathcal{D}\}}f(X_T^x)$ où $\mathcal{D}$ est un domaine ouvert de $\mathbb{R}^d$ et f une fonction dont le support est strictement inclus dans $\mathcal{D}$. L'auteur montre aussi que la version discrète du schéma d'Euler converge à la vitesse $\frac{1}{\sqrt{n}}$. Temam [121] s'est intéressé aux options asiatiques et a

obtenu une vitesse en $\frac{1}{n}$ pour des fonctionnelles du type $f\left(\int_0^T X_t^x dt\right)$ pour f une fonction lipschitzienne. Tanré [120] a montré que c'est également le cas pour des fonctionnelles du type $\int_0^T f(X_t^x)dt$ avec f seulement mesurable bornée. Citons également Seumen Tonou [113] qui s'est intéressé aux options *lookback* et qui a obtenu une vitesse en $\frac{1}{\sqrt{n}}$ pour la version discrète du schéma d'Euler.

Pour les fonctionnelles lipschitziennes, nous disposons d'un cadre mathématique adéquat pour formuler cette problématique : la distance de Wasserstein (on trouve dans certaines références d'autres terminologies pour cette distance comme la distance de Monge-Kantorovitch ou de Kantorovitch-Rubinstein). À ce sujet, et plus généralement au sujet du transport optimal, nous renvoyons le lecteur aux ouvrages de Villani [124] et de Rachev et Rüschendorf [101, 102]. En l'occurrence, grâce à la formule de dualité de Kantorovitch, la distance de Wasserstein entre P_{X^x} et $P_{X^{x,n}}$, les lois de $(X_t^x)_{t\in[0,T]}$ et de $(X_t^{x,n})_{t\in[0,T]}$ respectivement, s'écrit comme

$$d_W(P_{X^x}, P_{X^{x,n}}) = \sup_{\phi\in Lip_1} \left|\mathbb{E}\left(\phi((X_t^x)_{t\in[0,T]})\right) - \mathbb{E}\left(\phi((X_t^{x,n})_{t\in[0,T]})\right)\right|$$

où $Lip_1 = \left\{\phi : \mathcal{C}([0,T],\mathbb{R}^d) \to \mathbb{R}; \forall (x,y) \in \mathcal{C}([0,T],\mathbb{R}^d)^2, |\phi(x)-\phi(y)| \leq \sup_{t\in[0,T]} |x_t - y_t|\right\}$.

Contrôler la distance de Wasserstein entre la solution de l'EDS et son schéma d'Euler est certainement difficile. Nous apportons une première réponse en estimant la distance de Wasserstein entre les marginales de ces processus uniformément en temps. Plus précisément, nous montrons le résultat suivant :

Théorème 3 *Supposons que*
- $\forall 1 \leq i \leq d$ *et* $\forall 1 \leq j \leq r$, $\quad b_i, \sigma_{i,j} \in \mathcal{C}_b^\infty(\mathbb{R}^d)$.
- $\exists \eta > 0$ *tel que* $\forall x, \xi \in \mathbb{R}^d$, $\quad \xi^* a(x)\xi \geq \eta\|\xi\|^2$ *où* a *désigne la matrice* $\sigma\sigma^*$ *(on note la transposition par une étoile).*

Alors, il existe une constante $C > 0$ *indépendante de* n *tel que*

$$\sup_{0\leq t\leq T} d_W\left(P_{X_t^x}, P_{X_t^{x,n}}\right) \leq \frac{C}{n},$$

où, $\forall t \in [0,T], P_{X_t^x}$ *et* $P_{X_t^{x,n}}$ *désignent respectivement les lois de* X_t^x *et de* $X_t^{x,n}$.

Sous les mêmes hypothèses que ce théorème, Guyon [52] a obtenu un développement limité de la différence entre la densité de la solution et celle de son schéma d'Euler à tout instant. Le terme principal de ce développement explose pour les temps petits et ne permet pas de retrouver notre résultat. Récemment, et indépendamment de notre travail, Gobet et Labart [50] ont montré une majoration plus fine de la différence entre la densité de la solution et son schéma d'Euler, et ce pour des EDS inhomogènes en temps et sous des hypothèses plus faibles que celles de Guyon [52]. Nous montrons comment déduire notre théorème à partir de leur résultat et nous donnons une preuve directe basée sur une méthode probabiliste/analytique classique, à la différence de l'approche de Gobet et Labart [50] basée sur le calcul de Malliavin.

2. Modélisation de la dépendance en finance

La deuxième partie de cette thèse est composée de deux chapitres. Le premier est consacré à la modélisation jointe entre un indice boursier et les actions qui le composent et le deuxième traite

de la modélisation des risques de contrepartie dans un portefeuille de crédit. Bien qu'ils concernent deux domaines différents de la finance, en l'occurrence le marché actions et le risque de crédit, ces deux travaux partagent le même souci d'une meilleure modélisation de la dépendance. Dans un premier cas, c'est la dépendance entre les actions qui composent un même indice boursier qui nous intéresse et dans le deuxième, c'est la dépendance entre la qualité de signature des composants d'un portefeuille de crédit.

2.1 Un modèle couplant indice et actions

Un indice boursier est une collection d'actions, souvent représentative d'un marché global ou d'un secteur industriel particulier. Sa valeur est déterminée par une somme pondérée des prix des actions qui le composent, les poids étant typiquement proportionnels à la capitalisation boursière des composants de l'indice.

Bien que le marché des indices soit plus liquide que celui des actions individuelles, il existe relativement peu de travaux sur la modélisation des indices. La principale difficulté provient de la grande dimension des problématiques découlant de la modélisation jointe d'un indice et des actions qui le composent. De plus, plusieurs études empiriques mettent en évidence un comportement particulier pour la volatilité implicite de l'indice comparé à la volatilité des actions qui rend la modélisation encore plus difficile. En effet, on observe que le *smile*[2] de volatilité d'un indice est généralement plus pentu que celui d'une action ordinaire (voir par exemple Bakshi *et al.* [6], Bollen et Whaley [16], Branger et Schlag [18]).

Par conséquent, il est difficile d'avoir un modèle global qui permette de se caler à la fois sur les prix d'options sur indice et sur les prix d'options sur les actions qui le composent. L'approche standard consiste à prendre un modèle *smilé* pour chaque action, généralement un modèle à volatilité locale ou un modèle à volatilité stochastique, et à apposer une matrice de corrélation, généralement constante et estimée de manière historique puisqu'une estimation implicite est beaucoup plus délicate. On reconstruit alors la dynamique de l'indice à partir des dynamiques individuelles des actions. À ce titre, citons l'article de Avellaneda *et al.* [5] qui reconstruisent la volatilité locale de l'indice à partir des volatilités locales des actions en utilisant une technique basée sur des développements de grandes déviations. Aussi, Lee *et al.* [82] reconstruisent le développement de Gram-Charlier de la densité de probabilité de l'indice à partir des actions en utilisant une méthode des moments.

Dans le chapitre 4, nous proposons une approche nouvelle pour la modélisation jointe de l'indice et de ses composantes. Intuitivement, puisque l'indice synthétise le marché et représente les vues et les anticipations des acteurs financiers sur l'état de l'économie, il n'est pas déraisonnable de penser que l'évolution du prix d'un indice boursier influe sur les prix des actions. Sous cet angle de vue, l'indice n'est plus simplement une somme pondérée de prix mais devient un facteur qui agit sur ces mêmes prix. Plus précisément, nous postulons un cadre de modélisation dans lequel les volatilités de l'indice et des actions qui le composent sont reliées.

2. On appelle *smile* de volatilité la courbe qui donne la volatilité implicite en fonction du prix d'exercice. Contrairement au cadre offert par le modèle de Black & Scholes, cette courbe n'est pas constante, la terminologie *smile* provient de la forme ressemblant à un sourire qu'on observe sur certains marchés.

On note par I_t^M la valeur à l'instant t d'un indice composé de M actions :

$$I_t^M = \sum_{j=1}^{M} w_j S_t^{j,M}, \tag{12}$$

où $S_t^{j,M}$ représente la valeur de l'action j au temps t et les poids $(w_j)_{j=1\dots M}$ sont supposés constants.

Sous la probabilité risque-neutre, on spécifie les EDS suivantes pour l'évolutions des actions :

$$\forall j \in \{1, \dots, M\}, \quad \frac{dS_t^{j,M}}{S_t^{j,M}} = (r - \delta_j)dt + \beta_j\, \sigma(t, I_t^M)dB_t + \eta_j(t, S_t^{j,M})dW_t^j, \tag{13}$$

avec

- r le taux d'intérêt sans risque,
- $\delta_j \in [0, \infty[$ le taux de dividende continu de l'action j,
- β_j le coefficient beta habituel de l'action j qui relie les rendements de l'action aux rendements de l'indice (voir Sharpe [114]). Il est défini par $\frac{Cov(r_j, r_I)}{Var(r_I)}$ où r_j (respectivement r_I) est le taux de rendement de l'action j (respectivement de l'indice).
- $(B_t)_{t\in[0,T]}, (W_t^1)_{t\in[0,T]}, \dots, (W_t^M)_{t\in[0,T]}$ sont des mouvements Browniens indépendants.
- Les fonctions $\sigma, \eta_1, \dots, \eta_M$ vérifient les bonnes hypothèses qui assurent que le modèle est bien défini.

La dépendance entre les dynamiques des actions découle du terme de volatilité commun $\sigma(t, I_t^M)$. On peut voir notre modèle comme un modèle à un facteur. D'ailleurs, le pendant discret de ce modèle a été proposé par Cizeau *et al.* [22] qui montrent qu'un simple modèle à un facteur, non gaussien, permet de retrouver la structure de dépendance entre les actions, particulièrement dans des conditions extrêmes de marché (volatilité importante de l'indice).

En outre, les coefficients de corrélations entre actions sont stochastiques et dépendent à la fois des actions et de l'indice. En particulier, on vérifie que, comme il est communément observé sur les marché, plus l'indice est volatile, plus les coefficients de corrélation sont importants.

2.1.1 Un modèle simplifié

La plupart des indices sont composés d'un grand nombre d'actions. Par exemple, le CAC40 est composé de quarante actions, l'EUROSTOXX 50 et l'indice S&P500 en possèdent respectivement 50 et 500. Nous pouvons tirer profit de cette observation en regardant ce qui se passe quand M tend vers $+\infty$. Nous simplifions alors considérablement notre modèle. Plus précisément, considérons le l'EDS suivante

$$\begin{aligned} \forall j \in \{1, \dots, M\}, \quad \frac{dS_t^j}{S_t^j} &= (r - \delta_j)dt + \beta_j\, \sigma(t, I_t)dB_t + \eta_j(t, S_t^j)dW_t^j \\ \frac{dI_t}{I_t} &= (r - \delta_I)dt + \sigma(t, I_t)dB_t. \end{aligned} \tag{14}$$

Nous contrôlons les distances L^p entre $(I_t^M)_{t\in[0,T]}$ et $(I_t)_{t\in[0,T]}$ d'une part et entre $(S_t^{j,M})_{t\in[0,T]}$ et $(S_t^j)_{t\in[0,T]}$ d'autre part, pour j allant de 1 à M. Les estimations obtenues sont en pratique très faibles pour de grandes valeurs de M. Notre modèle initial peut donc être approché par ce modèle simplifié, dans lequel l'indice suit un modèle à volatilité locale et les actions individuelles suivent un

modèle à volatilité stochastique, composé d'une partie intrinsèque et d'une partie commune dirigée par l'indice. Afin d'éviter les opportunités d'arbitrages, il est aussi utile de considérer l'indice comme la somme pondérée des actions : $\overline{I}_t^M = \sum_{j=1}^M w_j S_t^j$. Nous contrôlons également la distance L^p entre $(I_t^M)_{t\in[0,T]}$ et $(\overline{I}_t^M)_{t\in[0,T]}$.

2.1.2 Calibration

La dernière partie de ce chapitre est consacré à la calibration des modèles proposés. La calibration, c'est à dire l'estimation des paramètres d'un modèle de manière à coller le plus possible aux prix observés sur le marché, représente un enjeu crucial en finance. Notre montrons comment calibrer le modèle simplifié à la fois pour l'indice et pour les actions qui le composent. Cette calibration simultanée et cohérente au sein d'un même modèle constitue le principal avantage de notre approche.

En fait, la calibration de l'indice dans le modèle simplifié revient à calibrer un modèle à volatilité locale, ce qui est un problème bien connu (voir Dupire [37]). En pratique, on postule une forme paramétrique pour la volatilité et on estime les paramètres par une méthode de moindre carrés. La calibration du coefficient de volatilité intrinsèque de l'action est plus ardue. Notons au passage que le fait d'avoir favorisé la calibration de l'indice par rapport à celle des actions est en ligne avec le marché puisque les options sur indice sont généralement plus traitées que les options sur actions.

Nous proposons une méthode de calibration originale pour l'action. Au lieu d'estimer le coefficient de volatilité intrinsèque, nous présentons une méthode pour simuler des trajectoires suivant la bonne loi, c'est à dire la loi qui permet de retrouver les prix d'options observés sur le marché. En effet, en se basant sur les résultats de Gyöngy [53], le coefficient η_j qui permet de retrouver les bons prix d'options peut s'exprimer en fonction de la volatilité locale et d'une espérance conditionnelle. On obtient alors une EDS non-linéaire au sens de McKean (voir Sznitman [116] ou Méléard [89] pour une introduction aux EDS non linéaires et à la propagation du chaos). Nous proposons une méthode d'estimation non-paramétrique de l'espérance conditionnelle et la simulation du système d'EDS, linéaires cette fois, qui en découle par un simple schéma d'Euler.

La fin du chapitre est consacré aux résultats numériques. En utilisant des jeux de données réels pour l'indice EUROSTOXX 50, nous observons que notre modèle simplifié permet de retrouver les courbes de volatilité implicite de l'indice et des actions qui le composent. Une parfaite calibration de notre modèle original est relativement compliquée mais grâce aux différents résultats sur l'erreur d'approximation en passant à la limite $M \to \infty$, il est raisonnable de faire une calibration sur le modèle simplifié et de l'utiliser dans le modèle original. D'ailleurs, nous illustrons numériquement la qualité de notre approximation en regardant les volatilités implicites obtenues avec les deux modèles. Enfin, la comparaison avec le modèle standard du marché qui consiste à prendre une matrice de corrélation constante met clairement en évidence les défauts de ce dernier : la structure de dépendance n'est pas assez flexible pour retrouver le bon *smile* de volatilité de l'indice. Par conséquent, les prix d'options sensibles à la corrélation entre actions (comme l'option worst-off considéré ici) sont plus fiables avec notre cadre de modélisation.

2.2 Estimation d'un modèle à intensité pour la gestion des risques. Extension à un modèle de *frailty* dynamique

Le dernier chapitre s'inscrit dans le cadre de ma collaboration avec le département des risques de la banque IXIS CIB, devenue aujourd'hui NATEXIS. Ce travail a donné lieu à un article, écrit avec Jean-David Fermanian et Martin Delloye, qui a été publié dans la revue *Risk*.

Avant d'aborder le vif du sujet, commençons par une petite introduction sur le risque de crédit. Par risque de crédit, on entend ici le risque lié à la défaillance d'une contrepartie, c'est à dire son incapacité à honorer ses engagements financiers.

L'exemple récent de la crise financière déclenchée par les *subprimes*, ces produits dérivées de crédit hautement risqués qui ont défrayé la chronique et causé des pertes colossales pour les particuliers, les institutions financières et même les états, est venu rappeler l'importance du risque de crédit. Par le passé, l'intérêt pour les banques de mieux prendre en compte ce risque était déjà justifié par l'expansion du marché des dérivées de crédit, par la recrudescence de chocs macro-économiques violents qui induisent des défauts en masse, comme les attentats du 11 septembre, l'éclatement de la bulle internet ou encore les épidémies qui ont secoué l'Asie, mais aussi par l'éclatement d'affaires de crédit retentissantes comme l'affaire Enron ou WorldCom ou encore la faillite de l'état argentin. S'ajoutent à cela les contraintes réglementaires, comme les directives européennes concernant les méthodologies de calcul du capital réglementaire (Bâle II), qui seront probablement durcies à la suite de la crise économique que nous traversons.

Dans ce contexte, les banques sont de plus en plus amenées à modéliser correctement le risque de crédit qu'elles encourent, soit dans une optique de pricing et de couverture de produits dérivées (CDS, CDO, ...), soit, et c'est ce qui nous intéresse en l'occurrence, dans une optique de gestion de risque et de contrôle d'activité : calcul de capital économique, allocation d'un capital adéquat à chaque branche d'activité, évaluation de la performance de ces branches au regard des risques encourus, diversification du risque et réduction de ce dernier par l'imposition de limites d'exposition ...
Pour ce faire, elles s'appuient sur des informations sur la santé financière et la capacité des entreprises à payer leurs dettes en temps et en heure. Ces informations sont l'apanage des agences de notation (Moody's, Standard&Poor's et Fitch pour citer les plus fameuses) qui sont des organismes indépendants censés donner une opinion objective[3] sur le risque de défaillance d'un émetteur ou d'une émission. Cette opinion est symbolisée par des notes (communément appelées *ratings*) qui permettent d'apprécier la qualité de crédit d'une entité et de la comparer à d'autres entités. Le système de notation diffère d'une agence de *rating* à une autre mais les lettres-symboles **AAA**, **AA**, **A**, **BBB**, **BB**, **B** et **CCC** sont devenues un langage international. Ces notes ont un impact très important pour les entités en question puisqu'elles ont pour conséquence une augmentation de leur *spread* de crédit, c'est à dire la différence entre le taux exigé par le marché et le taux sans risque. Par conséquent, le risque de crédit ne se limite pas au risque de défaut de la contrepartie mais il faut aussi prendre en compte ce qu'on appelle le risque de *downgrade*, c'est à dire la dégradation du *rating*. Le défaut n'est alors qu'un *rating* particulier. Il faut aussi garder à l'esprit que dans un portefeuille, les risques de crédit individuels sont fortement corrélés, notamment pour les entreprises relevant d'un même secteur industriel ou d'une même zone géographique. Il peut aussi y avoir des

3. Les agences de notation se sont retrouvées sous le feu des critiques pour leur rôle dans la crise des *subprimes*. Leur détracteurs les accusent de conflit d'intérêt et de manque d'objectivité les ayant amené à sous-estimer le risque de certains émetteurs et produits financiers émis sur le marché.

effets de contagion et des mécanismes de défauts en cascade qui induisent de grandes pertes [4]. D'où l'importance de l'enjeu d'une bonne modélisation de la dépendance entre les risques individuels dans un même portefeuille.

Les deux grandes familles de modèles de portefeuille de crédit sont les modèles structurels et les modèles à intensité, ou encore modèles à forme réduite. Décrivons rapidement en quoi consistent ces deux approches :

- **Modèle structurel**
  C'est un modèle économique qui considère que le risque de contrepartie est directement lié à la structure capitalistique de l'entreprise : il y a défaut quand la valeur des actifs de celle-ci passe au dessous d'une certaine valeur critique, représentative de la dette. Les nombreuses extensions de ce modèle diffèrent les unes des autres par la modélisation de la valeur des actifs et la modélisation de la valeur seuil. À titre d'exemple, le premier modèle de crédit dû à Merton [90] suppose que la valeur des actifs suit un processus log-normal sous la probabilité historique. Il y a défaut quand celle-ci est inférieure à la valeur de la dette.
  Cette modélisation présente plusieurs avantages : le cadre théorique est bien maîtrisé puisqu'il est proche de celui des options, l'interprétation des paramètres est aisée et la dépendance entre événements de défaut est simple à mettre en oeuvre (corrélation entre valeurs d'actifs). Toutefois, la valeur des actifs est inobservable et la calibration des paramètres est problématique.

- **Modèle à intensités**
  Contrairement à l'approche précédente, l'idée ici est de décrire directement la loi du défaut. En effet, on considère que le processus qui mène au défaut est un processus aléatoire pour lequel on peut définir une intensité de défaut qui s'interprète en première approximation comme la probabilité instantanée de faire défaut (Jarrow *et al.* [59], Duffie et Singleton [36]). La modélisation de cette intensité, notamment à l'aide de variables explicatives adéquates, permet de remonter jusqu'à la loi du défaut.
  Cette modélisation ne s'appuie pas sur des cadres théoriques restreints et permet une calibration facile sur des données observables (historiques de défauts ou plus généralement de transitions de *rating*). Il est néanmoins difficile de trouver de bonnes variables explicatives du défaut et la prise en compte de la dépendance entre événements de défaut n'est pas aisée.

Dans cette thèse, nous développons au dernier chapitre un modèle à intensités pour la mesure du risque de défaut qui permet de modéliser simultanément, et de manière consistante, toutes les transitions de *ratings* qui surviennent dans un grand portefeuille de crédit. Nous postulons que les transitions de *ratings* sont indépendantes conditionnellement à la valeur de certains facteurs macro-économiques explicites et facilement observables sur le marché. Plus précisément, en s'inspirant du modèle de Cox (voir Lando [79]), nous écrivons, pour une entreprise i du portefeuille de crédit, l'intensité de transition d'un *rating* h à un *rating* j à l'instant t comme

$$\alpha_{hji}(t|z) = \alpha_{hj0} \exp(\beta'_{hj} z_{hji}(t)), \tag{15}$$

où α_{hj0} est une constante, β_{hj} est un vecteur de paramètres inconnus et $z_{hji}(t)$ représente la valeur à l'instant t d'un vecteur regroupant des variables intrinsèques à l'entreprise i et des variables macro-économiques communes à toutes les entreprises, censées expliquer au mieux les transitions

4. La récente crise financière où on a vu les états agir très vite de peur du risque systémique illustre bien cet aspect.

de *rating*. La dépendance entre ces dernières est donc générée par les variables macro-économiques communes.

Contrairement à ce qu'on trouve dans la littérature, où l'objectif principal est le pricing et la couverture de produits dérivées de crédit, notre approche est purement historique et nous montrons comment calibrer aisément les paramètres de notre modèle sur la base de données de transitions de *ratings* passées uniquement, exemptes de l'interférence de l'aversion au risque des investisseurs sur les marchés. Nous travaillons donc exclusivement sous la probabilité historique et nous essayons de caler au mieux le modèle à partir des événements de défauts passés afin de pouvoir prédire correctement les risques dans le futur, après avoir modélisé l'évolution des facteurs macro-économiques choisis.

En particulier, nous montrons comment calibrer les paramètres du modèle par une procédure de maximisation de la vraisemblance. La base historique sur laquelle on s'appuie est celle fournie par Standard&Poor's (CreditPro) qui enregistre des milliers de transitions de *ratings* individuels de firmes dans le monde (principalement pour l'Amérique du nord et l'Europe) depuis 1981. Plusieurs résultats empiriques sont présentés et nous montrons que le modèle arrive à bien reproduire les transitions de *ratings* observées. Toutefois, sa principale faiblesse réside dans le faible niveau de dépendance qu'il génère entre les transitions de *ratings* au sein du portefeuille.

Notre contribution la plus importante a été de proposer une amélioration de ce modèle qui permet de mieux tenir compte de la dépendance entre transitions de *ratings*. En s'inspirant de littérature sur l'analyse de survie (voir par exemple Clayton et Cuzick [23] ou Hougaard [56]), nous introduisons des facteurs inobservables qui agissent de manière multiplicative sur les intensités de transition. Ces variables sont appelées "*dynamic frailties*" : "*dynamic*" parce qu'on considère des processus qui bougent dans le temps et non des variables statiques comme c'est souvent le cas dans la littérature, et "*frailty*", qui veut dire en anglais fragilité, parce que l'effet de ces variables est d'augmenter l'intensité de transition ou de défaut. Plus précisément, pour un groupe de transitions de *ratings* (par exemple les *downgrades* d'une note), on remplace la spécification (15) par

$$\alpha_{hji}(t|z) = \gamma_t \alpha_{hj0} \exp(\beta'_{hj} z_{hji}(t)), \tag{16}$$

où $(\gamma_t)_{t=1,\dots,T}$ est une chaîne de Markov définie par

$$\gamma_1 = \tilde{\gamma}_1, \quad \gamma_t = \gamma_{t-1}\tilde{\gamma}_t$$

avec $(\tilde{\gamma}_t)_{t=1,\dots,T}$ une suite de variables indépendantes et identiquement distribuées suivant une loi gamma de paramètre inconnu α, de telle sorte que $\mathbb{E}[\tilde{\gamma}_t] = 1$ et $Var(\tilde{\gamma}_t) = 1/\alpha$.

L'estimation de ce nouveau modèle est plus difficile que pour le modèle précédent, notamment à cause du caractère dynamique de la *frailty* : un modèle de *frailty* statique permet certes d'avoir une expression explicite de la vraisemblance mais ne réussit pas à atteindre des niveaux de dépendance convenables entre transitions de *ratings*. Pour calibrer notre modèle, nous proposons encore une estimation par maximisation de la vraisemblance mais cette fois via l'algorithme EM (E pour *expectation* et M pour *maximisation*). Cet algorithme est bien connu en statistique, surtout pour l'estimation par maximum de vraisemblance de modèles avec données manquantes (voir l'excellent livre de McLachlan et Krishnan [88] sur le sujet). En fait, pour être précis, nous utilisons un algorithme du type SEM (S pour *stochastic*, voir Celeux et Diebolt [21]) : dans l'étape *expectation*, nous calculons une espérance par une méthode MCMC (*Monte Carlo Markov Chain*, voir par exemple Robert et Casella [104]). Cette idée a déjà été proposée dans la littérature : voir Diebolt et Ip [32] par exemple.

Les résultats empiriques obtenus sont très satisfaisants. En particulier, nous montrons que le modèle *dynamic frailty* permet d'atteindre des niveaux de dépendance élevés entre les transitions de *ratings*, surtout à horizon lointain.

Notons enfin que, à l'époque de la réalisation de ce travail, l'idée d'introduire une *frailty* dynamique dans la modélisation du risque de crédit n'était pas très répandue. Parmi les rares travaux voisins, citons Metayer [91] qui introduit un modèle de *frailty* statique permettant un calcul explicite de la vraisemblance, Schönbucher [111] qui étudie l'effet de contagion entre firmes dont les *frailties* statiques sont fortement corrélées ou encore Koopman *et al.* [71] qui proposent une famille de modèles proches du notre mais étudient une méthode d'estimation différente, basée sur le filtre de Kalman. Depuis, l'idée d'introduire une variable inobservable dans les modèles de crédit a fait son chemin : à titre d'exemple, citons les travaux de Duffie *et al.* [35], Runggaldier et Frey [108], Runggaldier et Fontana [107] ou encore Giesecke et Azizpour [46].

Première partie

Méthodes de simulation exacte et schémas de discrétisation d'EDS. Applications en finance

Chapitre 1

Méthodes de Monte Carlo exactes et application au pricing d'options asiatiques

Ce chapitre correspond à un article écrit avec mon directeur de thèse Benjamin Jourdain (voir Jourdain et Sbai [60]). Il a été publié dans la revue *Monte Carlo Methods and Applications*.

Abstract. Taking advantage of the recent literature on exact simulation algorithms (Beskos *et al.* [13]) and unbiased estimation of the expectation of certain functional integrals (Wagner [126], Beskos *et al.* [14] and Fearnhead *et al.* [38]), we apply an exact simulation based technique for pricing continuous arithmetic average Asian options in the Black & Scholes framework. Unlike existing Monte Carlo methods, we are no longer prone to the discretization bias resulting from the approximation of continuous time processes through discrete sampling. Numerical results of simulation studies are presented and variance reduction problems are considered.

Introduction

Although the Black & Scholes framework is very simple, it is still a challenging task to efficiently price Asian options. Since we do not know explicitly the distribution of the arithmetic sum of log-normal variables, there is no closed form solution for the price of an Asian option. By the early nineties, many researchers attempted to address this problem and hence different approaches were studied including analytic approximations (see Turnball and Wakeman [122], Vorst [125], Levy [83] and more recently Lord [85]), PDE methods (see Vecer [123], Rogers and Shi [105], Ingersoll [58], Dubois and Lelievre [34]), Laplace transform inversion methods (see Geman and Yor [45], Geman and Eydeland [44]) and, of course, Monte Carlo simulation methods (see Kemna and Vorst [67], Broadie and Glasserman [19], Fu *et al.* [42]).

Monte Carlo simulation can be computationally expensive because of the usual statistical error. Variance reduction techniques are then essential to accelerate the convergence (one of the most efficient techniques is the Kemna&Vorst control variate based on the geometric average). One must also account for the inherent discretization bias resulting from approximating the continuous average of the stock price with a discrete one. It is crucial to choose with care the discretization scheme in order to have an accurate solution (see Lapeyre and Temam [81]). The main contribution of our work is to fully address this last feature by the use, after a suitable change of variables, of an exact simulation method inspired from the recent work of Beskos *et al.* [13, 14] and Fearnhead *et al.* [38].

In the first part of the paper, we recall the algorithm introduced by Beskos *et al.* [13] in order to simulate sample-paths of processes solving one-dimensional stochastic differential equations. By a suitable change of variables, one may suppose that the diffusion coefficient is equal to one. Then, according to the Girsanov theorem, one may deal with the drift coefficient by introducing an exponential martingale weight. Because of the one-dimensional setting, the stochastic integral in this exponential weight is equal to a standard integral with respect to the time variable up to the addition of a function of the terminal value of the path. Under suitable assumptions, conditionally on a Brownian path, an event with probability equal to the normalized exponential weight can be simulated using a Poisson point process. This allows to accept or reject this Brownian path as a path solution to the SDE with diffusion coefficient equal to one. In finance, one is interested in computing expectations rather than exact simulation of the paths. In this perspective, computation of the exponential importance sampling weight is enough. The entire series expansion of the exponential function permits to replace this exponential weight by a computable weight with the same conditional expectation given the Brownian path. This idea was first introduced by Wagner [126, 127, 128, 129] in a statistical physics context and it was very recently revisited by Beskos *et al.* [14] and Fearnhead *et al.* [38] for the estimation of partially observed diffusions. Some of the assumptions necessary to implement the exact algorithm of Beskos *et al.* [13] can then be weakened.

The second part is devoted to the application of these methods to option pricing within the Black & Scholes framework. Throughout the paper, $S_t = S_0 \exp\left(\sigma W_t + (r - \delta - \frac{\sigma^2}{2})t\right)$ represents the stock price at time t, T the maturity of the option, r the short interest rate, σ the volatility parameter, δ the dividend rate and $(W)_{t\in[0,T]}$ denotes a standard Brownian motion on the risk-neutral probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We are interested in computing the price $C_0 = \mathbb{E}\left(e^{-rT} f\left(\alpha S_T + \beta \int_0^T S_t dt\right)\right)$ of a European option with pay-off $f\left(\alpha S_T + \beta \int_0^T S_t dt\right)$ assumed to be square integrable under the risk neutral measure $\mathbb{P}$. The constants α and β are two

given non-negative parameters.

When $\alpha > 0$, we remark that, by a change of variables inspired by Rogers and Shi [105], $\alpha S_T + \beta \int_0^T S_t dt$ has the same law as the solution at time T of a well-chosen one-dimensional stochastic differential equation. Then it is easy to implement the exact methods previously presented. The case $\alpha = 0$ of standard Asian options is more intricate. The previous approach does not work and we propose a new change of variables which is singular at initial time. It is not possible to implement neither the exact simulation algorithm nor the method based on the unbiased estimator of Wagner [126] and we propose a pseudo-exact hybrid method which appears as an extension of the exact simulation algorithm. In both cases, one first replaces the integral with respect to the time variable in the function f by an integral with respect to time in the exponential function. Because of the nice properties of this last function, exact computation is possible.

1.1 Exact Simulation techniques

1.1.1 The exact simulation method of Beskos *et al.* [13]

In a recent paper, Beskos *et al.* [13] proposed an algorithm which allows to simulate exactly the solution of a 1-dimensional stochastic differential equation. Under some hypotheses, they manage to implement an acceptance-rejection algorithm over the whole path of the solution, based on recursive simulation of a biased Brownian motion. Let us briefly recall their methodology. We refer to [13] for the demonstrations and a detailed presentation.

Consider the stochastic process $(\xi_t)_{0\leq t\leq T}$ determined as the solution of a general stochastic differential equation of the form :

$$\left\{ \begin{array}{rcl} d\xi_t & = & b(\xi_t)dt + \sigma(\xi_t)dW_t \\ \xi_0 & = & \xi \in \mathbb{R} \end{array} \right. \tag{1.1}$$

where b and σ are scalar functions satisfying the usual Lipschitz and growth conditions with σ non vanishing. To simplify this equation, Beskos *et al.* [13] suggest to use the following change of variables : $X_t = \eta(\xi_t)$ where η is a primitive of $\frac{1}{\sigma}$ ($\eta(x) = \int_{\cdot}^{x} \frac{1}{\sigma(u)} du$).
Under the additional assumption that $\frac{1}{\sigma}$ is continuously differentiable, one can apply Itô's lemma to get

$$\begin{aligned} dX_t &= \eta'(\xi_t)d\xi_t + \frac{1}{2}\eta''(\xi_t)\, d< \xi, \xi >_t \\ &= \frac{b(\xi_t)}{\sigma(\xi_t)}dt + dW_t - \frac{\sigma'(\xi_t)}{2}dt \\ &= \underbrace{\left(\frac{b(\eta^{-1}(X_t))}{\sigma(\eta^{-1}(X_t))} - \frac{\sigma'(\eta^{-1}(X_t))}{2} \right)}_{a(X_t)} dt + dW_t \end{aligned}$$

So $\xi_t = \eta^{-1}(X_t)$ where $(X_t)_t$ is a solution of the stochastic differential equation

$$\left\{ \begin{array}{rcl} dX_t & = & a(X_t)dt + dW_t \\ X_0 & = & x. \end{array} \right. \tag{1.2}$$

Thus, without loss of generality, one can start from equation (1.2) instead of (1.1).

Let us denote by $(W_t^x)_{t\in[0,T]}$ the process $(W_t + x)_{t\in[0,T]}$, by $\mathbb{Q}_{W^x}$ its law and by $\mathbb{Q}_X$ the law of the process $(X_t)_{t\in[0,T]}$. From now on, we will denote by $(Y_t)_{t\in[0,T]}$ the canonical process, that is the coordinate mapping on the set $C([0,T],\mathbb{R})$ of real continuous maps on $[0,T]$ (see Revuz and Yor [103] or Karatzas and Shreve [63]).

One needs the following assumption to be true

<u>*Assumption 1*</u> : *Under $\mathbb{Q}_{W^x}$, the process*

$$L_t = \exp\left[\int_0^t a(Y_u)dY_u - \frac{1}{2}\int_0^t a^2(Y_u)du\right]$$

is a martingale.

According to Rydberg [109] (see the proof of Proposition 4 where we give his argument on a specific example), a sufficient condition for this assumption to hold is

-Existence and uniqueness in law of a solution to the SDE (1.2).

-$\forall t \in [0,T]$, $\displaystyle\int_0^t a^2(Y_u)du < \infty$, $\mathbb{Q}_X$ and $\mathbb{Q}_{W^x}$ almost surely on $C([0,T],\mathbb{R})$.

Thanks to this assumption, one can apply the Girsanov theorem to get that $\mathbb{Q}_X$ is absolutely continuous with respect to $\mathbb{Q}_{W^x}$ and its Radon-Nikodym derivative is equal to

$$\frac{d\mathbb{Q}_X}{d\mathbb{Q}_{W^x}} = \exp\left[\int_0^T a(Y_t)dY_t - \frac{1}{2}\int_0^T a^2(Y_t)dt\right].$$

Consider A the primitive of the drift a, and assume that

<u>*Assumption 2*</u> : *a is continuously differentiable.*

Since, by Itô's lemma, $A(W_T^x) = A(x) + \int_0^T a(W_t^x)dW_t^x + \frac{1}{2}\int_0^T a'(W_t^x)dt$, we have

$$\frac{d\mathbb{Q}_X}{d\mathbb{Q}_{W^x}} = \exp\left[A(Y_T) - A(x) - \frac{1}{2}\int_0^T a^2(Y_t) + a'(Y_t)dt\right].$$

Before setting up an acceptance-rejection algorithm using this Radon-Nikodym derivative, a last step is needed. To ensure the existence of a density $h(u)$ proportional to $\exp(A(u) - \frac{(u-x)^2}{2T})$, it is necessary and sufficient that the following assumption holds

<u>*Assumption 3*</u> : *The function $u \mapsto \exp(A(u) - \frac{(u-x)^2}{2T})$ is integrable.*

Finally, let us define a process Z_t distributed according to the following law $\mathbb{Q}_Z$

$$\mathbb{Q}_Z = \int_{\mathbb{R}} \mathcal{L}\Big((W_t^x)_{t\in[0,T]}|W_T^x = y\Big)h(y)dy$$

where the notation $\mathcal{L}(.|.)$ stands for the conditional law. One has

$$\frac{d\mathbb{Q}_X}{d\mathbb{Q}_Z} = \frac{d\mathbb{Q}_X}{d\mathbb{Q}_{W^x}}\frac{d\mathbb{Q}_{W^x}}{d\mathbb{Q}_Z} = C\exp\left[-\frac{1}{2}\int_0^T a^2(Y_t) + a'(Y_t)dt\right]$$

where C is a normalizing constant. At this level, Beskos *et al.* [13] need another assumption

Assumption 4 : *The function* $\phi : x \mapsto \frac{a^2(x)+a'(x)}{2}$ *is bounded from below.*

Therefore, one can find a lower bound k of this function and eventually the Radon-Nikodym derivative of the change of measure between X and Z takes the form

$$\frac{d\mathbb{Q}_X}{d\mathbb{Q}_Z} = Ce^{-kT} \exp\left[-\int_0^T \phi(Y_t) - k\,dt\right].$$

The idea behind the exact algorithm is the following : suppose that one is able to simulate a continuous path $Z_t(\omega)$ distributed according to $\mathbb{Q}_Z$ and let $M(\omega)$ be an upper bound of the mapping $t \mapsto \phi(Z_t(\omega)) - k$. Let N be an independent random variable which follows the Poisson distribution with parameter $TM(\omega)$ and let $(U_i, V_i)_{i=1\ldots N}$ be a sequence of independent random variables uniformly distributed on $[0,T] \times [0, M(\omega)]$. Then, the number of points (U_i, V_i) which fall below the graph $\{(t, \phi(Z_t(\omega)) - k); t \in [0,T]\}$ is equal to zero with probability $\exp\left[-\int_0^T \phi(Z_t(\omega)) - k\,dt\right]$. Actually, simulating the whole path $(Z_t)_{t\in[0,T]}$ is not necessary. It is sufficient to determine an upper bound for $\phi(Z_t) - k$ since, as pointed out by the authors, it is possible to simulate recursively a Brownian motion on a bounded time interval by first simulating its endpoint, then simulating its minimum or its maximum and finally simulating the other points [1]. For this reason, one needs the following assumption for the algorithm to be feasible :

Assumption 5 : *Either* $\limsup_{u\to+\infty} \phi(u) < +\infty$ *or* $\limsup_{u\to-\infty} \phi(u) < +\infty$.

Suppose for example that $\limsup_{u\to+\infty} \phi(u) < +\infty$. The exact algorithm of Bekos *et al.* [13] then takes the following form :

Algorithm 1

1. *Draw the ending point* Z_T *of the process* Z *with respect to the density* h.
2. *Simulate the minimum* m *of the process* Z *given* Z_T.
3. *Fix an upper bound* $M(m) = \sup\{\phi(u) - k; u \geq m\}$ *for the mapping* $t \mapsto \phi(Z_t) - k$.
4. *Draw* N *according to the Poisson distribution with parameter* $TM(m)$ *and draw* $(U_i, V_i)_{i=1\ldots N}$, *a sequence of independent variables uniformly distributed on* $[0,T] \times [0, M(m)]$.
5. *Fill in the path of* Z *at the remaining times* $(U_i)_{i=1\ldots N}$.
6. *Evaluate the number of points* $(V_i)_{i=1\ldots N}$ *such that* $V_i \leq \phi(Z_{U_i}) - k$.
   If it is equal to zero, then return the simulated path Z.
   Else, return to step 1.

This algorithm gives exact skeletons of the process X, solution of the SDE (1.2). Once accepted, a path can be further recursively simulated at additional times without any other acceptance/rejection criteria. We also point out that the same technique can be generalized by replacing the Brownian motion in the law of the proposal Z by any process that one is able to simulate recursively by first simulating its ending point, its minimum/maximum and then the other points. Also, the extension of the algorithm to the inhomogeneous case, where the drift coefficient a in (1.2), and therefore the function ϕ, depend on the time variable t, is straightforward given that the assumptions presented above are appropriately modified.

1. In their paper, the authors explain how to do such a decomposition of the Brownian path.

1.1.2 The unbiased estimator (U.E)

In finance, the pricing of contingent claims often comes down to the problem of computing an expectation of the form

$$C_0 = \mathbb{E}\left(f(X_T)\right) \tag{1.3}$$

where X is a solution of the SDE (1.2) and f is a scalar function such that $f(X_T)$ is square integrable. In a simulation based approach, one is usually unable to exhibit an explicit solution of this SDE and will therefore resort to numerical discretization schemes, such as the Euler or Milstein schemes, which introduce a bias. Of course, the exact algorithm presented above avoids this bias. Here, we are going to present a technique which permits to compute exactly the expectation (1.3) while assumptions 4 and 5 on the function $\frac{a^2+a'}{2}$ which appears in the Radon-Nikodym derivative are relaxed.

Using the previous results and notations, we get, under the assumptions 1 and 2, that

$$C_0 = \mathbb{E}\left(f(W_T^x)\exp\left[A(W_T^x) - A(x) - \frac{1}{2}\int_0^T a^2(W_t^x) + a'(W_t^x)dt\right]\right). \tag{1.4}$$

In order to implement an importance sampling method, let us introduce a positive density ρ on the real line and a process $(Z_t)_{t\in[0,T]}$ distributed according to the following law $\mathbb{Q}_Z$

$$\mathbb{Q}_Z = \int_{\mathbb{R}} \mathcal{L}\Big((W_t^x)_{t\in[0,T]}|W_T^x = y\Big)\rho(y)dy.$$

By (1.4), one has

$$C_0 = \mathbb{E}\left(\psi(Z_T)\exp\left[-\int_0^T \phi(Z_t)dt\right]\right) \tag{1.5}$$

where $\psi : z \mapsto f(z)\frac{e^{A(z)-A(x)-\frac{(z-x)^2}{2T}}}{\sqrt{2\pi}\rho(z)}$ and $\phi : z \mapsto \frac{a^2(z)+a'(z)}{2}$. We do not impose ρ to be equal to the density h of the previous section. It is a free parameter chosen in such a way that it reduces the variance of the simulation.

In his first paper, Wagner [126] constructs an unbiased estimator of the expectation (1.5) when ψ is a constant, $(Z_t)_{t\in[0,T]}$ is an $\mathbb{R}^d-$valued Markov process with known transition function and ϕ is a measurable function such that $\mathbb{E}\left(e^{\int_0^T|\phi(Z_t)|dt}\right) < +\infty$. His main idea is to expand the exponential term in a power series, then, using the transition function of the underlying Markov process and symmetry arguments, he constructs a signed measure ν on the space $\mathcal{Y} = \bigcup_{n=0}^{+\infty}([0,T]\times\mathbb{R}^d)^{n+1}$ such that the expectation at hand is equal to $\nu(\mathcal{Y})$. Consequently, any probability measure μ on Y that is absolutely continuous with respect to ν gives rise to an unbiased estimator ζ defined on $(\mathcal{Y},\mu)$ via $\zeta(y) = \frac{d\nu}{d\mu}(y)$. In practice, a suitable way to construct such an estimator is to use a Markov chain with an absorbing state. Wagner also discusses variance reduction techniques, specially importance sampling and a shift procedure consisting on adding a constant c to the integrand ϕ and then multiplying by the factor e^{-cT} in order to get the right expectation. Wagner [128] extends the class of unbiased estimators by perturbing the integrand ϕ by a suitably chosen function ϕ_0 and then using mixed integration formulas representation. Very recently, Beskos *et al.* [14] obtained a simplified unbiased estimator for (1.5), termed Poisson estimator, using Wagner's idea of expanding the exponential in a power series and his shift procedure. To be specific, the Poisson estimator

writes

$$\psi(Z_T)e^{c_pT-cT}\prod_{i=1}^{N}\frac{c-\phi(Z_{V_i})}{c_P} \tag{1.6}$$

where N is a Poisson random variable with parameter c_P and $(V_i)_i$ is a sequence of independent random variables uniformly distributed on $[0,T]$. Fearnhead *et al.* [38] generalized this estimator allowing c and c_P to depend on Z and N to be distributed according to any positive probability distribution on $\mathbb{N}$. They termed the new estimator the generalized Poisson estimator. We introduce a new degree of freedom by allowing the sequence $(V_i)_i$ to be distributed according to any positive density on $[0,T]$. This gives rise to the following unbiased estimator for (1.5) :

Lemma 1 — *Let p_Z and q_Z denote respectively a positive probability measure on $\mathbb{N}$ and a positive probability density on $[0,T]$. Let N be distributed according to p_Z and $(V_i)_{i\in\mathbb{N}^*}$ be a sequence of independent random variables identically distributed according to the density q_Z, both independent from each other conditionally on the process $(Z_t)_{t\in[0,T]}$. Let c_Z be a real number which may depend on Z. Assume that*

$$\mathbb{E}\left(|\psi(Z_T)|e^{-c_ZT}\exp\left[\int_0^T|c_Z-\phi(Z_t)|dt\right]\right)<\infty.$$

Then

$$\psi(Z_T)e^{-c_ZT}\frac{1}{p_Z(N)\,N!}\prod_{i=1}^{N}\frac{c_Z-\phi(Z_{V_i})}{q_Z(V_i)} \tag{1.7}$$

is an unbiased estimator of C_0.

Proof : The result follows from

$$\begin{aligned}\mathbb{E}\left(\psi(Z_T)e^{-c_ZT}\frac{1}{p_Z(N)\,N!}\prod_{i=1}^{N}\frac{c_Z-\phi(Z_{V_i})}{q_Z(V_i)}\Big|(Z_t)_{t\in[0,T]}\right)&=\psi(Z_T)e^{-c_ZT}\sum_{n=0}^{+\infty}\frac{\left(\int_0^T c_Z-\phi(Z_t)dt\right)^n}{p_Z(n)\,n!}p_Z(n)\\&=\psi(Z_T)\exp\left(-\int_0^T\phi(Z_t)dt\right).\end{aligned}$$

□

Using (1.7), one is now able to compute the expectation at hand by a simple Monte Carlo simulation. The practical choice of p_Z and q_Z conditionally on Z is studied in the appendix 1.4.1.

As pointed out in Fearnhead *et al.* [38], this method is an extension of the exact algorithm method since, under assumptions 3, 4 and 5, the reinforced integrability assumption of Lemma 1 is always satisfied.

Indeed, suppose for example that $\limsup_{u\to+\infty}\phi(u)<+\infty$ and let k be a lower bound of ϕ, m_Z be the minimum of the process Z and M_Z an upper bound of $\{\phi(u)-k, u\geq m_Z\}$. Then, taking

$c_Z = M_Z + k$ in Lemma 1 ensures the integrability condition :

$$\begin{aligned}\mathbb{E}\left(|\psi(Z_T)|e^{-(M_Z+k)T}e^{\int_0^T|M_Z+k-\phi(Z_t)|dt}\right) &= \mathbb{E}\left(|\psi(Z_T)|e^{-(M_Z+k)T}e^{\int_0^T M_Z+k-\phi(Z_t)dt}\right)\\ &= \mathbb{E}\left(|\psi(Z_T)|e^{-\int_0^T\phi(Z_t)dt}\right) < \infty\end{aligned}$$

and hence, one is allowed to write that

$$C_0 = \mathbb{E}\left(\psi(Z_T)e^{-(M_Z+k)T}\frac{1}{p_Z(N)N!}\prod_{i=1}^N\frac{M_Z+k-\phi(Z_{V_i})}{q_Z(V_i)}\right).$$

Better still, the random variable $\psi(Z_T)e^{-(M_Z+k)T}\frac{1}{p_Z(N)N!}\prod_{i=1}^N\frac{M_Z+k-\phi(Z_{V_i})}{q_Z(V_i)}$ is square integrable when p_Z is the Poisson distribution with parameter M_ZT+k and q_Z is the uniform distribution on $[0,T]$ since we have then

$$\begin{aligned}\mathbb{E}\left(\left(\psi(Z_T)e^{-(M_Z+k)T}\frac{1}{p_Z(N)N!}\prod_{i=1}^N\frac{M_Z+k-\phi(Z_{V_i})}{q_Z(V_i)}\right)^2\right) &= \mathbb{E}\left(\psi^2(Z_T)\prod_{i=1}^N\left(1-\frac{\phi(Z_{V_i})}{M_Z+k}\right)^2\right)\\ &\leq \mathbb{E}\left(\psi^2(Z_T)\right) < \infty.\end{aligned}$$

The last inequality follows from the square integrability of f : whenever one is able to simulate from the density h, introduced in the exact algorithm, by doing rejection sampling, there exists a density ρ such that ψ, which is equal to $f(Z_T)\frac{h(Z_T)}{\rho(Z_T)}$ up to a constant factor, is dominated by f and so is square integrable.

The square integrability property is very important in that we use a Monte Carlo method. We see that, whenever the exact algorithm is feasible, the unbiased estimator of lemma 1 is a simulable square integrable random variable, at least for the previous choice of p_Z and q_Z.

Remark 2 — *One can derive two estimators of C_0 from the result of Lemma 1 :*

$$\begin{aligned}\delta_1 &= \frac{1}{n}\sum_{i=1}^n f(Z_T^i)\frac{e^{A(Z_T^i)-A(x)-\frac{(Z_T^i-x)^2}{2T}}}{\sqrt{2\pi}\rho(Z_T^i)}e^{-c_ZT}\frac{1}{p_Z(N^i)\,N^i!}\prod_{j=1}^{N^i}\frac{c_Z-\phi(Z^i_{V_j^i})}{q_Z(V_j^i)}\\ \delta_2 &= \frac{\displaystyle\sum_{i=1}^n f(Z_T^i)\frac{e^{A(Z_T^i)-A(x)-\frac{(Z_T^i-x)^2}{2T}}}{\sqrt{2\pi}\rho(Z_T^i)}\frac{1}{p_Z(N^i)\,N^i!}\prod_{j=1}^{N^i}\frac{c_Z-\phi(Z^i_{V_j^i})}{q_Z(V_j^i)}}{\displaystyle\sum_{i=1}^n\frac{e^{A(Z_T^i)-A(x)-\frac{(Z_T^i-x)^2}{2T}}}{\sqrt{2\pi}\rho(Z_T^i)}\frac{1}{p_Z(N^i)\,N^i!}\prod_{j=1}^{N^i}\frac{c_Z-\phi(Z^i_{V_j^i})}{q_Z(V_j^i)}}.\end{aligned}$$

1.2 Application : the pricing of continuous Asian options

In the Black & Scholes model, the stock price is the solution of the following SDE under the risk-neutral measure $\mathbb{P}$

$$\frac{dS_t}{S_t} = (r-\delta)dt + \sigma dW_t \tag{1.8}$$

where all the parameters are constant : r is the short interest rate, δ is the dividend rate and σ is the volatility.
Throughout, we denote $\gamma = r - \delta - \frac{\sigma^2}{2}$. The path-wise unique solution of (1.8) is

$$S_t = S_0 \exp(\sigma W_t + \gamma t)\,.$$

We consider an option with pay-off of the form

$$f\left(\alpha S_T + \beta \int_0^T S_t dt\right) \tag{1.9}$$

where f is a given function such that $\mathbb{E}\left(f^2\left(\alpha S_T + \beta \int_0^T S_t dt\right)\right) < \infty$, T is the maturity of the option and α, β are two given non negative parameters [2]. Note that for $\alpha = 0$, this is the pay-off of a standard continuous Asian option.

The fundamental theorem of arbitrage-free pricing ensures that the price of the option under consideration is

$$C_0 = \mathbb{E}\left(e^{-rT} f\left(\alpha S_T + \beta \int_0^T S_u du\right)\right).$$

At first sight, the problem seems to involve two variables : the stock price and the integral of the stock price with respect to time. Dealing with the PDE associated with Asian option pricing, Rogers and Rogers and Shi [105] used a suitable change of variables to reduce the spatial dimension of the problem to one. We are going to use a similar idea.

Let

$$\xi_t = \left(\alpha S_0 + \beta S_0 \int_0^t e^{-\sigma W_u - \gamma u} du\right) e^{\sigma W_t + \gamma t}.$$

We have that

$$\begin{aligned} \xi_t &= \alpha S_0 e^{\sigma W_t + \gamma t} + \beta S_0 \int_0^t e^{\sigma(W_t - W_u) + \gamma(t-u)} du \\ &= \alpha S_0 e^{\sigma B_t + \gamma t} + \beta S_0 \int_0^t e^{\sigma B_s + \gamma s} ds \end{aligned}$$

where we set $B_s = W_t - W_{t-s}, \forall s \in [0,t]$. Clearly, $(B_s)_{s\in[0,t]}$ is a Brownian motion and thus the following lemma holds

Lemma 3 — *$\forall t \in [0,T]$, ξ_t and $\alpha S_t + \beta \int_0^t S_u du$ have the same law.*

As a consequence

$$C_0 = \mathbb{E}\left(e^{-rT} f(\xi_T)\right).$$

By applying Itô's lemma, we verify that the process $(\xi_t)_{t\geq 0}$ is a positive solution of the following 1-dimensional stochastic differential equation for which path-wise uniqueness holds

$$\left\{ \begin{array}{rcl} d\xi_t &=& \beta S_0 dt + \xi_t(\sigma dW_t + (\gamma + \frac{\sigma^2}{2})dt) \\ \xi_0 &=& \alpha S_0. \end{array} \right. \tag{1.10}$$

2. The underlying of this option is a weighted average of the stock price at maturity and the running average of the stock price until maturity with respective weights α and βT.

We are thus able to value C_0 by Monte Carlo simulation without resorting to discretization schemes using one of the exact simulation techniques described in the previous section. In the case $\alpha = 0$, one has to deal with the fact that ξ_t starts from zero which is the reason why we distinguish two cases.

1.2.1 The case $\alpha \neq 0$

We are going to apply both the exact algorithm of Beskos *et al.* [13] and the method based on the unbiased estimator of lemma 1.

We make the following change of variables to have a diffusion coefficient equal to 1 :

$$X_t = \frac{\log(\xi_t)}{\sigma} \Rightarrow \begin{cases} dX_t &= (\frac{\gamma}{\sigma} + \frac{\beta S_0}{\sigma} e^{-\sigma X_t})dt + dW_t \\ X_0 &= x \quad \text{with } x = \frac{\log(\alpha S_0)}{\sigma}. \end{cases} \tag{1.11}$$

Thus

$$C_0 = \mathbb{E}\left(e^{-rT} f(e^{\sigma X_T})\right).$$

The following proposition ensures that assumption 1 is satisfied.

Proposition 4 — *The process $(L_t)_{t\in[0,T]}$ defined by*

$$L_t = \exp\left[\int_0^T (\frac{\gamma}{\sigma} + \frac{\beta S_0}{\sigma} e^{-\sigma Y_t})\, dY_t - \frac{1}{2}\int_0^T (\frac{\gamma}{\sigma} + \frac{\beta S_0}{\sigma} e^{-\sigma Y_t})^2 dt\right]$$

is a martingale under $\mathbb{Q}_{W^x}$.

Proof : Under $\mathbb{Q}_{W^x}$, $(L_t)_{t\in[0,T]}$ is clearly a non-negative local martingale and hence a supermartingale. Then, it is a true martingale if and only if $\mathbb{E}_{\mathbb{Q}_{W^x}}(L_T) = 1$.

Checking the classical Novikov's or Kamazaki's criteria is not straightforward. Instead, we are going to use the approach developed by Rydberg [109] (see also Wong and Heyde [134]) who takes advantage of the link between explosions of SDEs and the martingale property of stochastic exponentials.

Let us define the following stopping times :

$$\tau_n(Y) = \inf\left\{t \in \mathbb{R}^+ \text{ such that } \int_0^t \left(\frac{\gamma}{\sigma} + \frac{\beta S_0}{\sigma} e^{-\sigma Y_u}\right)^2 du \geq n\right\},$$

with the convention $\inf\{\emptyset\} = +\infty$.

The stopped process $(L_{t\wedge\tau_n(Y)})_{t\in[0,T]}$ is a true martingale under $\mathbb{Q}_{W^x}$ since Novikov's condition is fulfilled. According to the Girsanov theorem, one can define a new probability measure $\mathbb{Q}_X^n$, which is absolutely continuous with respect to $\mathbb{Q}_{W^x}$, by its Radon-Nikodym derivative

$$\frac{d\mathbb{Q}_X^n}{d\mathbb{Q}_{W^x}} = L_{T\wedge\tau_n(Y)}.$$

Hence

$$\mathbb{E}_{\mathbb{Q}_X^n}\left(\mathbb{1}_{\{\tau_n(Y)>T\}}\right) = \mathbb{E}_{\mathbb{Q}_{W^x}}\left(\mathbb{1}_{\{\tau_n(Y)>T\}} L_{T\wedge\tau_n(Y)}\right).$$

Since $(\tau_n(Y))_{n\in\mathbb{N}}$ is a non decreasing sequence, we can pass to the limit in the right hand side We get

$$\lim_{n\to+\infty} \mathbb{Q}_X^n\left(\tau_n(Y) > T\right) = \mathbb{E}_{\mathbb{Q}_{W^x}}\left(\mathbb{1}_{\{\tau_\infty(Y)>T\}} L_{T\wedge\tau_\infty(Y)}\right)$$

where $\tau_\infty(Y)$ denotes the limit of the non decreasing sequence $(\tau_n(Y))_{n\in\mathbb{N}}$.

Under $\mathbb{Q}_{W^x}$, $(Y_t)_{t\in[0,T]}$ has the same law as a Brownian motion starting from x so $\tau_\infty(Y) = +\infty$, $\mathbb{Q}_{W^x}$ almost surely, and consequently

$$\mathbb{E}_{\mathbb{Q}_{W^x}}\left(L_T\right) = \lim_{n\to+\infty} \mathbb{Q}_X^n\left(\tau_n(Y) > T\right).$$

On the other hand, the Girsanov theorem implies that, under $\mathbb{Q}_X^n$, $(Y_t)_{t\in[0,T\wedge\tau_n(Y)]}$ solves a SDE of the form (1.11). To conclude the proof, it is sufficient to check that trajectorial uniqueness holds for this SDE. Indeed, the law of $(Y_t)_{t\in[0,T\wedge\tau_n(Y)]}$ under $\mathbb{Q}_X^n$ is the same as the law of $(Y_t)_{t\in[0,T\wedge\tau_n(Y)]}$ under $\mathbb{Q}_X$. Hence

$$\mathbb{Q}_X^n\left(\tau_n(Y) > T\right) = \mathbb{Q}_X\left(\tau_n(Y) > T\right) \underset{n\to+\infty}{\longrightarrow} \mathbb{Q}_X\left(\tau_\infty(Y) > T\right).$$

Clearly, $\int_0^t \left(\frac{\gamma}{\sigma} + \frac{\beta S_0}{\sigma} e^{-\sigma Y_u}\right)^2 du < \infty$, $\mathbb{Q}_X$ almost surely, so

$$\mathbb{E}_{\mathbb{Q}_{W^x}}\left(L_T\right) = \mathbb{Q}_X\left(\tau_\infty(Y) > T\right) = 1$$

as required.

In order to check trajectorial uniqueness for the SDE (1.11), we consider two solutions X^1 and X^2. We have that

$$d(X_t^1 - X_t^2) = \frac{\beta S_0}{\sigma}\left(e^{-\sigma X_t^1} - e^{-\sigma X_t^2}\right) dt \Rightarrow d|X_t^1 - X_t^2| = \frac{\beta S_0}{\sigma} sign(X_t^1 - X_t^2)\left(e^{-\sigma X_t^1} - e^{-\sigma X_t^2}\right) dt.$$

So

$$|X_t^1 - X_t^2| = \frac{\beta S_0}{\sigma}\int_0^t sign(X_s^1 - X_s^2)\left(e^{-\sigma X_t^1} - e^{-\sigma X_t^2}\right) ds \leq 0.$$

The last inequality follows from the fact that $x \mapsto e^{-\sigma x}$ is a decreasing function. Finally, almost surely, $\forall t \geq 0$, $X_t^1 = X_t^2$ which leads to strong uniqueness. □

Consequently, thanks to the Girsanov theorem, we have

$$\frac{d\mathbb{Q}_X}{d\mathbb{Q}_{W^x}} = \exp\left[\int_0^T \underbrace{(\frac{\gamma}{\sigma} + \frac{\beta S_0}{\sigma} e^{-\sigma Y_t})}_{a(Y_t)} dY_t - \frac{1}{2}\int_0^T (\frac{\gamma}{\sigma} + \frac{\beta S_0}{\sigma} e^{-\sigma Y_t})^2 dt\right]. \tag{1.12}$$

Set $A(u) = \int_0^u a(x)dx = \frac{\gamma}{\sigma}u + \frac{\beta S_0}{\sigma^2}(1 - e^{-\sigma u})$. Then

$$\frac{d\mathbb{Q}_X}{d\mathbb{Q}_{W^x}} = \exp\left[A(Y_T) - A(x) - \frac{1}{2}\int_0^T a^2(Y_t) + a'(Y_t)dt\right].$$

The function $u \mapsto \exp\left(A(u) - \frac{(u-Y_0)^2}{2T}\right) = \exp\left(\frac{\gamma}{\sigma}u + \frac{\beta S_0}{\sigma^2}(1-e^{-\sigma u}) - \frac{(u-Y_0)^2}{2T}\right)$ is clearly integrable so we can define a new process $(Z_t)_{t\in[0,T]}$ distributed according to the following law $\mathbb{Q}_Z$

$$\mathbb{Q}_Z = \int_{\mathbb{R}} \mathcal{L}\Big((W_t)_{t\in[0,T]}|W_T = y\Big)h(y)dy$$

where the probability density h is of the form

$$h(u) = C\exp\left(A(u) - \frac{(u-Y_0)^2}{2T}\right) \qquad \text{with } C \text{ a normalizing constant.} \tag{1.13}$$

Remark 5 — *Simulating from this probability distribution is not difficult (see the appendix 1.4.2 for an appropriate method of acceptance/rejection sampling).*

We have

$$\frac{d\mathbb{Q}_X}{d\mathbb{Q}_Z} = C\exp\left[-\int_0^T \frac{1}{2}(a^2(Y_t) + a'(Y_t))dt\right].$$

Set $\phi(x) = \frac{a^2(x)+a'(x)}{2} = \frac{(\frac{\gamma}{\sigma}+\frac{\beta S_0}{\sigma}e^{-\sigma x})^2 - \beta S_0 e^{-\sigma x}}{2}$. A direct calculation gives

$$\inf_{x\in\mathbb{R}} \phi(x) = \begin{cases} \frac{\gamma^2}{2\sigma^2} & \text{if } 2\gamma \geq \sigma^2 \\ \phi\left(\frac{1}{\sigma}\log(\frac{2\beta S_0}{\sigma^2-2\gamma})\right) & \text{otherwise.} \end{cases}$$

Set $k = \inf_{x\in\mathbb{R}} \phi(x)$. Finally, we get

$$\frac{d\mathbb{Q}_X}{d\mathbb{Q}_Z} = Ce^{-kT}\exp\left[-\int_0^T \phi(Y_t) - k\,dt\right].$$

We check that

$$\lim_{x\to+\infty} \phi(x) = \frac{\gamma^2}{2\sigma^2} < \infty$$
$$\lim_{x\to-\infty} \phi(x) = +\infty.$$

Hence we can apply the algorithm 1 to simulate exactly X_T and compute $C_0 = \mathbb{E}\left(e^{-rT}f(e^{\sigma X_T})\right)$ by Monte Carlo. On the other hand, using (1.12) we get

$$C_0 = \mathbb{E}\left(e^{-rT}f(e^{\sigma W_T^x})\exp\left[A(W_T^x) - A(x) - \frac{1}{2}\int_0^T a^2(W_t^x) + a'(W_t^x)dt\right]\right)$$

and we can also use the unbiased estimator presented in the previous section to compute this expectation.

Remark 6 — *We also applied the exact algorithm based on a geometric Brownian motion instead of the standard Brownian motion which seems more intuitive given the form of the SDE (1.10). The algorithm is feasible because we can simulate recursively a drifted Brownian motion and therefore a geometric Brownian motion by an exponential change of variables. The results we obtained were not different from the first method.*

Numerical computation

For numerical tests, we consider the case

$$f(x) = (x - K)_+$$

which corresponds to the European call option with strike K. Using the exact simulation algorithm presented above, we can simulate the underlying $\alpha S_T + \beta \int_0^T S_t dt$ at maturity (see Figure 1.1). Then, all we have to do is a simple Monte Carlo method to get the price of the option under consideration. Using the unbiased estimator, we get

$$C_0 = \mathbb{E}\left(e^{-rT}(e^{\sigma Z_T} - K)_+ \frac{e^{A(Z_T) - A(x) - \frac{(Z_T - x)^2}{2T}}}{\sqrt{2\pi}\rho(Z_T)} e^{-(M_Z + k)T} \frac{1}{p_Z(N)N!} \prod_{i=1}^{N} \frac{M_Z + k - \phi(Z_{V_i})}{q(V_i)} \right)$$

where $(Z_t)_{t\in[0,T]}, \rho, M_Z, k, p_Z$ and q_Z are defined as in section 1.1.2. In order to ensure square integrability, we choose p_Z to be a Poisson distribution with parameter $M_Z T + k$ and q_Z to be the uniform distribution on $[0, T]$. For the density ρ, a good choice is to consider the density that we use to simulate from the distribution h by rejection sampling.

We test these exact methods against a standard discretization scheme with the variance reduction technique of Kemna and Vorst [67]. As pointed out by Lapeyre and Temam [81], the discretization of the integral by a simple Riemannian sum is not efficient. Instead, we use the trapezoidal discretization. In the sequel, we will denote this method by Trap+KV. The table 1.1 gives the results we obtained for the following arbitrary set of parameters : $S_0 = 100$, $K = 100$, $r = 0.05$, $\sigma = 0.3$, $\delta = 0$, $T = 1$, $\alpha = 0.6$ and $\beta = 0.4$. The computation has been made on a computer with a 2.8 Ghz Intel Penthium 4 processor. We intentionally choose a large number of simulations in order to show the influence of the number of time steps when using a discretization scheme.

Method	M	N	Acceptance rate	Price	C.I at 95%	CPU
Trap+KV	10	10^6	-	11.46	$[11.43, 11.48]$	5 s
	20			11.46	$[11.43, 11.49]$	9 s
	50			11.47	$[11.44, 11.5]$	21 s
Exact Simulation	-	10^6	24%	11.46	$[11.43, 11.5]$	81 s
U.E ($c_P = M_Z, c_Z = M_Z + k$)	-	10^6	-	11.46	$[11.43, 11.49]$	17 s
U.E ($c_P = c_Z = 1/T$)	-	10^6	-	11.46	$[11.43, 11.49]$	6 s

Table 1.1: Price of the option (1.9) using a standard discretization technique and exact simulation methods.

Empirical evidence shows that the exact simulation method is quite slow. This is mainly due to the fact that the rejection algorithm has a little acceptance rate (24% according to table 1.1). Using a geometric Brownian motion instead of a standard Brownian motion did not improve the results. Also, simulating recursively a Brownian path conditionally on its terminal value and its minimum is time consuming.

The unbiased estimator is more efficient, especially when we can avoid the recursive simulation of the Brownian path. To do so, we choose for p_Z a Poisson distribution with mean $c_P T$ where c_P

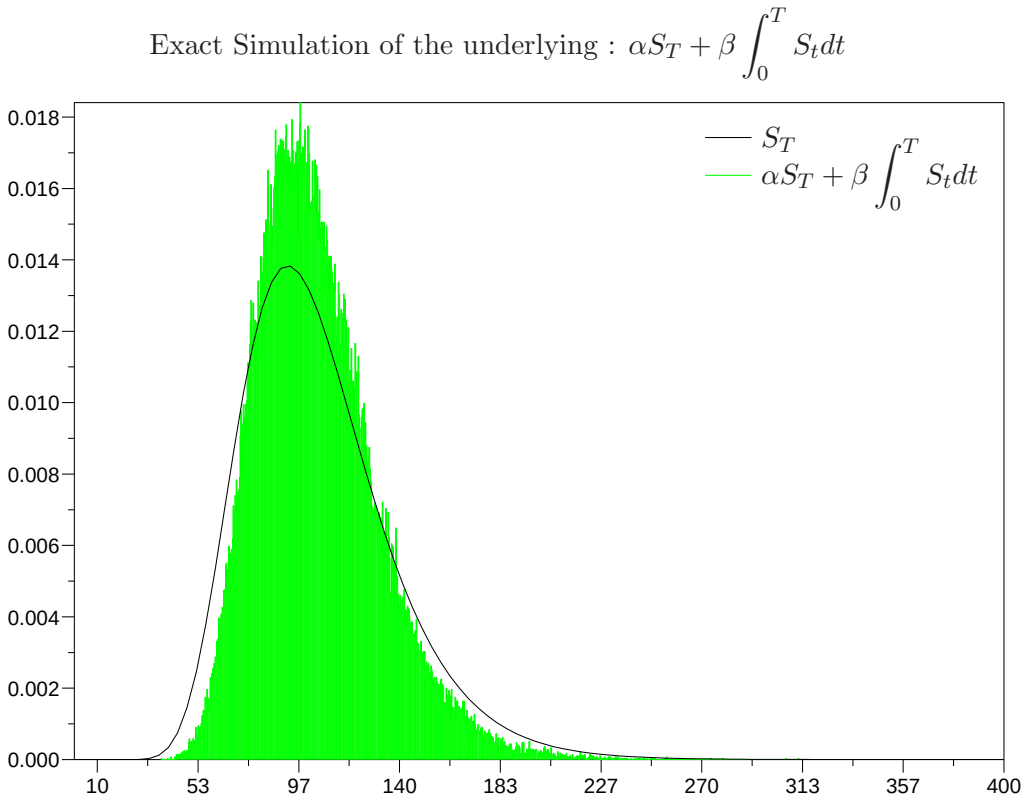

Figure 1.1: Histogram of 10^5 independent realizations of $\alpha S_T + \beta \int_0^T S_t dt$ for $\alpha = 0.6$ and $\beta = 0.4$ compared with the lognormal distribution of S_T.

is a free parameter. If we assume that the integrability condition in lemma 1 holds, then we can write that

$$C_0 = \mathbb{E}\left(e^{-rT}(e^{\sigma Z_T} - K)_+ \frac{e^{A(Z_T)-A(x)-\frac{(Z_T-x)^2}{2T}}}{\sqrt{2\pi}\rho(Z_T)} e^{c_P T - c_Z T} \prod_{i=1}^{N} \frac{c_Z - \phi(Z_{V_i})}{c_P} \right).$$

Regarding the dependence of the exact simulation method with respect to the parameters α and β, it is intuitive that whenever $\alpha >> \beta$, the method performs well since the logarithm of the underlying is not far from the logarithm of the geometric Brownian motion on which we do rejection-sampling. The table 1.2 confirms this intuition. We see that we cannot apply the algorithm for small values of α and then let $\alpha \to 0$ to treat the case $\alpha = 0$.

1.2.2 Standard Asian options : the case $\alpha = 0$ and $\beta > 0$

A standard Asian option is a European option on the average of the stock price over a determined period until maturity. An Asian call, for example, has a pay-off of the form $(\frac{1}{T}\int_0^T S_u du - K)_+$. With our previous notations, it corresponds to the case $\alpha = 0$, $\beta = \frac{1}{T}$ and $f(x) = (x - K)_+$.

$\frac{\alpha}{\alpha+\beta}$	0.3	0.4	0.5	0.6	0.7
Acceptance Rate	0.003%	0.47%	5.66%	24.43%	53.85%

Table 1.2: Influence of the parameter $\frac{\alpha}{\alpha+\beta}$ on the acceptance rate of the exact algorithm.

The change of variables we used above is no longer suitable because it starts from zero when $\alpha = 0$. Instead, we consider the following new definition of the process ξ

$$\left\{ \begin{array}{rcl} \xi_t & = & \frac{S_0}{t} \int_0^t e^{\sigma(W_t - W_u) + \gamma(t-u)} du \\ \xi_0 & = & S_0. \end{array} \right. \tag{1.14}$$

Obviously, the two variables ξ_T and $\frac{1}{T}\int_0^T S_u du$ have the same law. Hence, the price of the Asian option becomes

$$C_0 = \mathbb{E}\left(e^{-rT} f\left(\frac{1}{T}\int_0^T S_u du\right)\right) = \mathbb{E}\left(e^{-rT} f(\xi_T)\right).$$

Remark 7 — *The pricing of floating strike Asian options is also straightforward using this method. It is even more natural to consider these options since it unveils the appropriate change of variables as we shall see below.*

Let us consider a floating strike Asian call for example. We have to compute

$$C_0 = \mathbb{E}\left(e^{-rT}\big(\frac{1}{T}\int_0^T S_u du - S_T\big)_+\right).$$

Using $\widetilde{S}_t = S_t e^{\delta t}$ as a numéraire (see the seminal paper of Geman et al. [43]), we immediately obtain that

$$C_0 = \mathbb{E}_{\mathbb{P}_{\widetilde{S}}}\left(S_0 e^{-\delta T}\big(\frac{1}{T}\int_0^T \frac{S_u}{S_T} du - 1\big)_+\right)$$

where $\mathbb{P}_{\widetilde{S}}$ is the probability measure associated to the numéraire $\widetilde{S}_t$. It is defined by its Radon-Nikodym derivative $\frac{d\mathbb{P}_{\widetilde{S}}}{d\mathbb{P}} = e^{\sigma W_T - \frac{\sigma^2}{2}T}$.

Under $\mathbb{P}_{\widetilde{S}}$, the process $B_t = W_t - \sigma t$ is a Brownian motion and we can write that

$$\begin{array}{rcl} C_0 & = & \mathbb{E}_{\mathbb{P}_{\widetilde{S}}}\left(S_0 e^{-\delta T}\big(\frac{1}{T}\int_0^T e^{\sigma(B_u - B_T) + (r-\delta+\frac{\sigma^2}{2})(u-T)} du - 1\big)_+\right) \\ & = & \mathbb{E}\left(S_0 e^{-\delta T}\big(\frac{1}{T}\int_0^T e^{\sigma(W_u - W_T) + (r-\delta+\frac{\sigma^2}{2})(u-T)} du - 1\big)_+\right) \\ & = & \mathbb{E}\left(e^{-\delta T}\big(\xi_T - S_0\big)_+\right) \end{array}$$

where ξ_t is the process defined by (1.14) but with $\gamma = r - \delta + \frac{\sigma^2}{2}$. We see therefore that the problem simplifies to the fixed strike Asian pricing problem.

Let us write down the stochastic differential equation that rules the process $(\xi_t)_{t\in[0,T]}$. Using Itô's lemma, we get

$$\left\{\begin{array}{rcl} d\xi_t & = & \frac{\xi_0-\xi_t}{t}dt + \xi_t\left(\sigma dW_t + (\gamma+\frac{\sigma^2}{2})dt\right) \\ \xi_0 & = & S_0. \end{array}\right.$$

Note that we are faced with a singularity problem near 0 because of the term $\frac{\xi_0-\xi_t}{t}$. We are going to reduce its effect using another change of variables.

Using Itô's lemma, we show that

$$C_0 = \mathbb{E}\left(e^{-rT} f\left(S_0 e^{X_T}\right)\right) \tag{1.15}$$

where $X_t = \log(\xi_t/\xi_0)$ solves the following SDE

$$\left\{\begin{array}{rcl} dX_t & = & \sigma dW_t + \gamma dt + \frac{e^{-X_t}-1}{t}dt \\ X_0 & = & 0. \end{array}\right. \tag{1.16}$$

Lemma 8 — *Existence and strong uniqueness hold for the stochastic differential equation (1.16).*

Proof : Existence is obvious since we have a particular solution X_t. The diffusion coefficient being constant and the drift coefficient being a decreasing function in the spatial variable, we have also strong uniqueness for the SDE (see the proof of Proposition 4). □

Because of the singularity of the term $\frac{e^{-X_t}-1}{t}$ in the drift coefficient, the law of $(X_t)_{t\geq 0}$ is not absolutely continuous with respect to the law of $(\sigma W_t)_{t\geq 0}$. That is why we now define $(Z_t)_{t\geq 0}$ by the following SDE with an affine inhomogeneous drift coefficient :

$$\left\{\begin{array}{rcl} dZ_t & = & \sigma dW_t + \gamma dt - \dfrac{Z_t}{t}dt \\ Z_0 & = & X_0 = 0. \end{array}\right. \tag{1.17}$$

The drift coefficient exhibits the same behavior as the one in (1.16) in the limit $t \to 0$ in order to ensure the desired absolute continuity property. It is affine in the spatial variable so that $(Z_t)_{t\geq 0}$ is a Gaussian process and as such is easy to simulate recursively.

Lemma 9 — *The process*

$$Z_t = \frac{\sigma}{t}\int_0^t s dW_s + \frac{\gamma}{2}t \tag{1.18}$$

is the unique solution of the stochastic differential equation (1.17).

Proof : Using Itô's Lemma, we easily check that Z_t given by (1.18) is a solution of (1.17). Again, constant diffusion coefficient and decreasing drift coefficient ensures strong uniqueness. □

Remark 10 — *For the computation of the price* $C_0 = \mathbb{E}\left(e^{-rT}(S_0 e^{X_T} - K)_+\right)$ *of a standard Asian call option, the random variable* $e^{-rT}(S_0 e^{Z_T} - K)_+$ *provides a natural control variate. Indeed, since* Z_T *is a Gaussian random variable with mean* $\frac{\gamma}{2}T$ *and variance* $\frac{\sigma^2 T}{3}$*, one has*

$$\mathbb{E}\left(e^{-rT}(S_0 e^{Z_T} - K)_+\right) = S_0 e^{(\frac{\gamma}{2}+\frac{\sigma^2}{6}-r)T}\mathcal{N}\left(d + \sigma\sqrt{\frac{1}{3}T}\right) - Ke^{-rT}\mathcal{N}(d)$$

where $\mathcal{N}$ *is the cumulative standard normal distribution function and* $d = \frac{\log(S_0/K)+\frac{\gamma}{2}T}{\sigma\sqrt{\frac{1}{3}T}}$.

Notice that in Kemna and Vorst [67], the authors suggest the use of the control variate $e^{-rT}\left(S_0 \exp\left(\frac{1}{T}\int_0^T \sigma W_t + \gamma t\, dt\right) - K\right)_+$ *which has the same law than* $e^{-rT}\left(S_0 e^{Z_T} - K\right)_+$ *as* $\frac{1}{T}\int_0^T \sigma W_t + \gamma t\, dt$ *is also a Gaussian variable with mean* $\frac{\gamma}{2}T$ *and variance* $\frac{\sigma^2 T}{3}$.

In order to define a new probability measure under which $(Z_t)_{t\geq 0}$ solves the SDE (1.16), one introduces

$$L_t = \exp\left[\int_0^t \frac{e^{-Z_s} - 1 + Z_s}{\sigma s} dW_s - \frac{1}{2}\int_0^t \left(\frac{e^{-Z_s} - 1 + Z_s}{\sigma s}\right)^2 ds\right].$$

Because of the singularity of the coefficients in the neighborhood of $s = 0$, one has to check that the integrals in L_t are well defined. This relies on the following lemma

Lemma 11 — *Let* $\epsilon > 0$*. In a random neighborhood of* $s = 0$*, we have*

$$|Z_s| \leq c s^{\frac{1}{2}-\epsilon} \text{ and } |X_s| \leq c s^{\frac{1}{2}-\epsilon}$$

where c *is a constant depending on* σ,γ *and* ϵ.

Since $\forall \epsilon > 0$,

$$\forall z \leq c s^{\frac{1}{2}-\epsilon}, \left(\frac{e^{-z} - 1 + z}{\sigma s}\right)^2 \leq C s^{-4\epsilon},$$

we can choose $\epsilon < \frac{1}{4}$ to deduce that L_t is well defined.

Proof : We easily check that the Gaussian process $(B_t)_{t\in[0,T]}$ defined by $B_t = \int_0^{(3t)^{\frac{1}{3}}} s dW_s$ is a standard Brownian motion. Thanks to the law of iterated logarithm for the Brownian motion (see for example Karatzas and Shreve [63] p. 112), there exists $t_1(\omega)$ such that [3],

$$\forall t \leq t_1(\omega), |B_t(\omega)| \leq t^{\frac{1}{2}-\frac{\epsilon}{3}}.$$

Therefore,

$$\forall t \leq (3t_1(\omega))^{\frac{1}{3}}, \quad |Z_t(\omega)| = \left|\frac{\sigma}{t} B_{\frac{t^3}{3}}(\omega) + \frac{\gamma}{2}t\right| \leq \frac{\sigma}{3^{\frac{1}{2}-\frac{\epsilon}{3}}} t^{\frac{1}{2}-\epsilon} + \frac{\gamma}{2}t.$$

3. ω is an element of the underlying probability space Ω.

Taking $c = \max(\frac{\sigma}{3^{\frac{1}{2}-\frac{\epsilon}{3}}}, \frac{\gamma}{2})$ yields

$$\forall t \leq (3t_1(\omega))^{\frac{1}{3}} \wedge 1, \quad |Z_t(\omega)| \leq ct^{\frac{1}{2}-\epsilon}.$$

On the other hand, recall that $X_t = \log(\xi_t/\xi_0) = \log\left(\frac{1}{t}e^{\sigma W_t+\gamma t}\int_0^t e^{-\sigma W_u-\gamma u}du\right)$. So, using the law of iterated logarithm for the Brownian motion, we deduce that there exists $t_2(\omega)$ such that

$$\forall t \leq t_2(\omega), \quad 0 \leq \frac{1}{t}e^{\sigma W_t(\omega)+\gamma t}\int_0^t e^{-\sigma W_u(\omega)-\gamma u}du \leq \frac{1}{t}e^{\sigma t^{\frac{1}{2}-\epsilon}+\gamma t}\int_0^t e^{\sigma u^{\frac{1}{2}-\epsilon}-\gamma u}du.$$

Denote $g(t) = \frac{1}{t}e^{\sigma t^{\frac{1}{2}-\epsilon}+\gamma t}\int_0^t e^{\sigma u^{\frac{1}{2}-\epsilon}-\gamma u}du$ and let us investigate the order in time near zero of this function. We have that

$$\begin{aligned} e^{\sigma t^{\frac{1}{2}-\epsilon}+\gamma t} &= 1+\sigma t^{\frac{1}{2}-\epsilon}+\mathcal{O}(t^{1-2\epsilon}) \\ \int_0^t e^{\sigma u^{\frac{1}{2}-\epsilon}-\gamma u}du &= t+\frac{\sigma}{\frac{3}{2}-\epsilon}t^{\frac{3}{2}-\epsilon}+\mathcal{O}(t^{2-2\epsilon}) \end{aligned}$$

hence

$$g(t) = 1+(\sigma+\frac{\sigma}{\frac{3}{2}-\epsilon})t^{\frac{1}{2}-\epsilon}+\mathcal{O}(t^{1-2\epsilon}),$$

so $X_t(\omega) \leq \log(g(t)) \underset{t\to 0}{\sim} (\sigma+\frac{\sigma}{\frac{3}{2}-\epsilon})t^{\frac{1}{2}-\epsilon}$, which ends the proof for X_t. □

Proposition 12 — *$(L_t)_{t\in[0,T]}$ is a martingale and, consequently, for all $g : \mathcal{C}([0,T]) \to \mathbb{R}$ measurable, the random variables $g((X_t)_{0\leq t\leq T})$ and $g((Z_t)_{0\leq t\leq T})L_T$ are simultaneously integrable and then*

$$\mathbb{E}\Big(g((X_t)_{0\leq t\leq T})\Big) = \mathbb{E}\Big(g((Z_t)_{0\leq t\leq T})L_T\Big).$$

Proof : The proof is similar to the proof of Proposition 4.

We have already shown existence and strong uniqueness for both SDE (1.16) and (1.17). Showing that the stopping time

$$\tau_n(Y) = \inf\left\{t \in \mathbb{R}^+ \text{ such that } \int_0^t \left(\frac{e^{-Y_s}-1+Y_s}{\sigma s}\right)^2 ds \geq n\right\}, \text{ with the convention } \inf\{\emptyset\} = +\infty,$$

have infinite limits when n tends to $+\infty$, $\mathbb{Q}_X$ and $\mathbb{Q}_Z$ almost surely, follows from the previous lemma.

□

One has

$$L_T = \exp\left[\int_0^T \frac{e^{-Z_t} - 1 + Z_t}{\sigma^2 t} dZ_t - \int_0^T \frac{e^{-Z_t} - 1 + Z_t}{\sigma^2 t}\left(\frac{e^{-Z_t} - 1 + Z_t}{2t} + \gamma - \frac{Z_t}{t}\right) dt\right].$$

Set $A(t,z) = \dfrac{1 - z + \frac{z^2}{2} - e^{-z}}{\sigma^2 t}$. The function $A :]0,T] \times \mathbb{R} \to \mathbb{R}$ is continuously differentiable in time and twice continuously differentiable in space. So, we can apply Itô's Lemma on the interval $[\epsilon, T]$ for $\epsilon > 0$:

$$A(T, Z_T) = A(\epsilon, Z_\epsilon) + \int_\epsilon^T \frac{e^{-Z_t} - 1 + Z_t}{\sigma^2 t} dZ_t - \int_\epsilon^T \frac{1 - Z_t + \frac{Z_t^2}{2} - e^{-Z_t}}{\sigma^2 t^2} dt + \int_\epsilon^T \frac{1 - e^{-Z_t}}{2t} dt$$

Using the lemma 9, we let $\epsilon \to 0$ to obtain

$$A(T, Z_T) = \int_0^T \frac{e^{-Z_t} - 1 + Z_t}{\sigma^2 t} dZ_t - \int_0^T \frac{1 - Z_t + \frac{Z_t^2}{2} - e^{-Z_t}}{\sigma^2 t^2} dt + \int_0^T \frac{1 - e^{-Z_t}}{2t} dt.$$

Then

$$L_T = \exp\left[A(T, Z_T) - \int_0^T \phi(t, Z_t) dt\right]$$

where ϕ is the mapping

$$\phi(t,z) = \frac{e^{-z} - 1 + z - \frac{z^2}{2}}{\sigma^2 t^2} + \frac{1 - e^{-z}}{2t} + \frac{e^{-z} - 1 + z}{\sigma^2 t}\left(\frac{e^{-z} - 1 + z}{2t} + \gamma - \frac{z}{t}\right). \tag{1.19}$$

By (1.15) and Proposition 12, we get

$$C_0 = \mathbb{E}\left(e^{-rT} f(S_0 e^{Z_T}) \exp\left[A(T, Z_T) - \int_0^T \phi(t, Z_t) dt\right]\right). \tag{1.20}$$

Since for each $t > 0$, $\lim_{z \to -\infty} \phi(t,z) = +\infty$ and $\lim_{z \to +\infty} \phi(t,z) = -\infty$, it is not possible to apply the exact algorithm. One can use the unbiased estimator, at least theoretically, if there exists a random variable c_Z measurable with respect to Z such that

$$\mathbb{E}\left(e^{A(T,Z_T) - (r + c_Z)T} |f(S_0 e^{Z_T})| e^{\int_0^T |c_Z - \phi(t, Z_t)| dt}\right) < \infty.$$

Unfortunately, this reinforced integrability condition is never satisfied :

Lemma 13 — *Assume that f is a non identically zero function. Let p_Z and q_Z denote respectively a positive probability measure on $\mathbb{N}$ and a positive probability density on $[0,T]$. Let N be distributed according to p_Z and $(U_i)_{i \in \mathbb{N}^*}$ be a sequence of independent random variables identically distributed according to the density q_Z, both independent conditionally on the process $(Z_t)_{t \in [0,T]}$. Then the random variable*

$$e^{A(T,Z_T) - rT} f(S_0 e^{Z_T}) \frac{1}{p_Z(N)\, N!} \prod_{i=1}^N \frac{-\phi(U_i, Z_{U_i})}{q_Z(U_i)} \tag{1.21}$$

is non integrable.

Proof : By conditioning on Z, one has

$$\begin{aligned}\Delta := \mathbb{E}\left(\frac{e^{A(T,Z_T)-rT}|f(S_0e^{Z_T})|}{p_Z(N)\,N!}\prod_{i=1}^N\frac{|\phi(U_i,Z_{U_i})|}{q_Z(U_i)}\right) &= \mathbb{E}\left(e^{A(T,Z_T)-rT}|f(S_0e^{Z_T})|e^{\int_0^T|\phi(t,Z_t)|dt}\right)\\ &\geq \mathbb{E}\left(e^{A(T,Z_T)-rT}|f(S_0e^{Z_T})|e^{\int_{\frac{T}{2}}^T|\phi(t,Z_t)|dt}\right)\end{aligned}$$

One can easily show that, $\forall z<0$ and $\forall t\in[\frac{T}{2},T]$, $\phi(t,z)\geq\overline{\phi}(z)$ where

$$\overline{\phi}(z)=\frac{e^{-z}-1+z-\frac{z^2}{2}}{\sigma^2(\frac{T}{2})^2}+\frac{e^{-z}-1+z}{\sigma^2\frac{T}{2}}\left(\frac{e^{-z}-1+z}{T}+\gamma^+-2\frac{z}{T}\right)$$

Since $\overline{\phi}(z)\underset{-\infty}{\sim}2\dfrac{e^{-2z}}{\sigma^2T^2}$, there exists $c<0$ such that for all $z<c$, $\overline{\phi}(z)\geq\frac{e^{-2z}}{\sigma^2T^2}$. Hence,

$$\begin{aligned}\Delta &\geq \mathbb{E}\left(e^{A(T,Z_T)-rT}|f(S_0e^{Z_T})|e^{\frac{1}{\sigma^2T^2}\int_{\frac{T}{2}}^T e^{-2Z_t}\mathbb{1}_{\{Z_t<c\}}dt}\right)\\ &\geq \mathbb{E}\left(e^{A(T,Z_T)-rT}|f(S_0e^{Z_T})|e^{-\frac{e^{-2c}}{2\sigma^2T}}e^{\frac{1}{\sigma^2T^2}\int_{\frac{T}{2}}^T e^{-2Z_t}dt}\right)\end{aligned}$$

Using Jensen's inequality we get

$$\Delta \geq \mathbb{E}\left(e^{A(T,Z_T)-rT}|f(S_0e^{Z_T})|e^{-\frac{e^{-2c}}{2\sigma^2T}}\exp\left(\frac{1}{2\sigma^2T}e^{-\frac{4}{T}\int_{\frac{T}{2}}^T Z_tdt}\right)\right)$$

We have seen in the proof of lemma 11 that $Z_t=\frac{\sigma}{t}B_{\frac{t^3}{3}}+\frac{\gamma}{2}t$ where $(B_t)_{t\geq0}$ is a standard Brownian motion. So, conditionally on Z_T, $\int_{\frac{T}{2}}^T Z_tdt$ is a gaussian random variable and hence $\Delta=+\infty$.

□

We are in a situation where $e^{A(T,Z_T)-rT}|f(S_0e^{Z_T})|\,\mathbb{E}\left[\left|\frac{1}{p_Z(N)\,N!}\prod_{i=1}^N\frac{-\phi(U_i,Z_{U_i})}{q_Z(U_i)}\right|\Big|(Z_t)_{t\in[0,T]}\right]$ is non integrable while $e^{A(T,Z_T)-rT}|f(S_0e^{Z_T})|\left|\mathbb{E}\left[\frac{1}{p_Z(N)\,N!}\prod_{i=1}^N\frac{-\phi(U_i,Z_{U_i})}{q_Z(U_i)}\Big|(Z_t)_{t\in[0,T]}\right]\right|$ is integrable since $\mathbb{E}\left(e^{-rT}|f(S_0e^{Z_T})|\exp\left[A(T,Z_T)-\int_0^T\phi(t,Z_t)dt\right]\right)<\infty$. Then, a natural idea would consist in considering, for a given $n\in\mathbb{N}^*$, the random variable

$$e^{A(T,Z_T)-rT}|f(S_0e^{Z_T})|\,\mathbb{E}\left[\left|\frac{1}{n}\sum_{j=1}^n\frac{1}{p_Z(N_j)\,N_j!}\prod_{i=1}^{N_j}\frac{-\phi(U_i^j,Z_{U_i^j})}{q_Z(U_i^j)}\right|\Big|(Z_t)_{t\in[0,T]}\right]$$

where $(N_j)_{1\leq j\leq n}$ are independent variables having the same law as N and $\left((U_i^j)_{i\in\mathbb{N}^*}\right)_{1\leq j\leq n}$ are independent sequences having the same law as $(U_i)_{i\in\mathbb{N}^*}$, both independent conditionally on the process $(Z_t)_{t\in[0,T]}$. The following general result tells us that this is not sufficient to circumvent integrability problems.

Lemma 14 — *Let Y and Z be two real random variables and $g:\mathbb{R}\to\mathbb{R}$ a given measurable function. Assume that $g(Z)\mathbb{E}(Y|Z)$ is integrable while $g(Z)\mathbb{E}(|Y||Z)$ is non integrable. Then, when $(Y_i)_{1\leq i\leq n}$ is a sequence of independent random variables having the same law as Y, $\forall n\in\mathbb{N}^*$, the random variable $g(Z)\mathbb{E}\left(|\frac{1}{n}\sum_{i=1}^nY_i||Z\right)$ is non integrable.*

Proof : Denote by e, e_1 and e_n three functions satisfying

$$\forall z \in \mathbb{R}, \quad e(z) = \mathbb{E}\left(Y|Z=z\right), \quad e_1(z) = \mathbb{E}\left(|Y_1|\,|Z=z\right) \quad \text{and } e_n(z) = \mathbb{E}\left(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i\right| |Z=z\right)$$

On the one hand, since $\int_{\mathbb{R}} |g(z)|\,|e(z)|\mathbb{P}_Z(dz) < \infty$ and $\int_{\mathbb{R}} |g(z)|\,e_1(z)\mathbb{P}_Z(dz) = +\infty$, where $\mathbb{P}_Z$ is the law of Z, we have that $\int_{\mathbb{R}} |g(z)|\,e_1(z)\mathbb{1}_{\{e_1(z)\geq 2|e(z)|\}}\mathbb{P}_Z(dz) = +\infty$.
On the other hand, $\forall z \in \mathbb{R}$,

$$\begin{aligned}
e_n(z) &\geq \frac{1}{n}\left[\mathbb{E}\left(\left|\sum_{i=1}^{n} Y_i\right| \mathbb{1}_{\{\forall 2\leq j\leq n, Y_j\geq 0\}}|Z=z\right) + \mathbb{E}\left(\left|\sum_{i=1}^{n} Y_i\right| \mathbb{1}_{\{\forall 2\leq j\leq n, Y_j<0\}}|Z=z\right)\right] \\
&\geq \frac{1}{n}\left[\mathbb{E}\left(Y_1^+|Z=z\right)\mathbb{P}\left(Y_1\geq 0|Z=z\right)^{n-1} + \mathbb{E}\left(Y_1^-|Z=z\right)\mathbb{P}\left(Y_1<0|Z=z\right)^{n-1}\right] \\
&= \frac{1}{n}\left[\frac{e_1(z)+e(z)}{2}\mathbb{P}\left(Y_1\geq 0|Z=z\right)^{n-1} + \frac{e_1(z)-e(z)}{2}\mathbb{P}\left(Y_1<0|Z=z\right)^{n-1}\right] \\
&\geq \frac{1}{n}\left[\frac{e_1(z)}{4}\mathbb{1}_{\{e_1(z)\geq 2|e(z)|\}}\mathbb{P}\left(Y_1\geq 0|Z=z\right)^{n-1} + \frac{e_1(z)}{4}\mathbb{1}_{\{e_1(z)\geq 2|e(z)|\}}\mathbb{P}\left(Y_1<0|Z=z\right)^{n-1}\right] \\
&\geq \frac{e_1(z)}{n2^n}\mathbb{1}_{\{e_1(z)\geq 2|e(z)|\}}
\end{aligned}$$

Hence, $\mathbb{E}\left[g(Z)\mathbb{E}\left(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i\right| |Z\right)\right] = \int_{\mathbb{R}} |g(z)|e_n(z)\mathbb{P}_Z(dz) = +\infty$. □

There is still hope yet. In the proof of Lemma 13, we saw that integrability problems appear when Z_t takes large negative values so that $\phi(t, Z_t)$ tends rapidly towards $+\infty$. Since $\lim_{z\to+\infty} \phi(t,z) = -\infty$, one possible issue is to split the function $\phi(t, Z_t)$ into a positive part and a negative part. The first term can be handled by the exact simulation technique whereas the second term, which as we shall see in the following section presents no integrability problems, can be handled by the unbiased estimator technique.

An hybrid pseudo-exact method

We rewrite (1.20) in the following form

$$C_0 = \mathbb{E}\left(e^{A(T,Z_T)-rT} f(S_0 e^{Z_T}) e^{\int_0^T \phi^-(t,Z_t)dt} e^{-\int_0^T \phi^+(t,Z_t)dt}\right). \tag{1.22}$$

Let p_Z and q_Z denote respectively a positive probability measure on $\mathbb{N}$ and a positive probability density on $[0,T]$. Let N be distributed according to p_Z and $(U_i)_{i\in\mathbb{N}^*}$ be a sequence of independent random variables identically distributed according to the density q_Z, both independent conditionally on the process $(Z_t)_{t\in[0,T]}$. Note that, since $e^{A(T,Z_T)-rT} f(S_0 e^{Z_T}) e^{\int_0^T |\phi^-(t,Z_t)|dt} e^{-\int_0^T \phi^+(t,Z_t)dt} = e^{A(T,Z_T)-rT} f(S_0 e^{Z_T}) e^{-\int_0^T \phi(t,Z_t)dt}$ is integrable, one has

$$C_0 = \mathbb{E}\left(e^{A(T,Z_T)-rT} f(S_0 e^{Z_T}) \frac{1}{p_Z(N)N!}\left(\prod_{i=1}^{N} \frac{\phi^-(U_i, Z_{U_i})}{q_Z(U_i)}\right) e^{-\int_0^T \phi^+(t,Z_t)dt}\right). \tag{1.23}$$

Remark 15 — *There is no hope that this estimator is square integrable. Indeed, one can show as in Lemma 13 that* $\mathbb{E}\left(e^{\int_0^T \left(\phi^-(t,Z_t)\right)^2 dt}\right) = +\infty$ *since* $(\phi^-(t,z))^2$ *is of order* z^4 *for large positive* z.

The idea then is to apply the exact simulation technique to simulate an event with probability $e^{-\int_0^T \phi^+(t,Z_t)dt}$. Since for each $t > 0$, $\lim_{z\to-\infty} \phi^+(t,z) = +\infty$, one needs to bound from above $\phi^+(t,z)$, uniformly with respect to $t \in [0,T]$, for $z > c$ where $c < 0$ is a given constant. Thanks to the following lemma, it is possible to do so but only uniformly with respect to $t \in [\epsilon, T]$ for all $\epsilon > 0$:

Lemma 16 — *For all* $0 < t \leq T$,

$$\sup_{z\geq 0} \phi^+(t,z) \leq \frac{\gamma^2}{\sigma^2} + \frac{\gamma}{\sigma^2 t} + \frac{1}{t}\left(\frac{1}{2} - \frac{\gamma}{\sigma^2}\right)^+$$

and

$$\forall c < 0,\ \sup_{z\in[c,0]} \phi^+(t,z) \leq \frac{e^{-c} - 1 + c}{\sigma^2 t^2}(1 + \gamma^+ t) + \frac{(e^{-c}-1)^2}{2\sigma^2 t^2} - \frac{c^2}{\sigma^2 t^2}.$$

Proof : Let $z > 0$. It is useful to distinguish two cases according to the sign of γ :

1. $\underline{\gamma \geq 0}$
   We rewrite ϕ in the following form

$$\phi(t,z) = \frac{e^{-z} - 1 + z - \frac{z^2}{2}}{\sigma^2 t^2} + \frac{1 - e^{-z}}{t}\left(\frac{1}{2} - \frac{\gamma}{\sigma^2}\right) + \frac{\gamma z}{\sigma^2 t} - \frac{z^2 - (z\wedge 1)^2}{2\sigma^2 t^2} + \frac{(e^{-z}-1)^2 - (z\wedge 1)^2}{2\sigma^2 t^2}$$

   First note that $\frac{e^{-z}-1+z-\frac{z^2}{2}}{\sigma^2 t^2} \leq 0$, $\frac{1-e^{-z}}{t}\left(\frac{1}{2} - \frac{\gamma}{\sigma^2}\right) \leq \frac{1}{t}\left(\frac{1}{2} - \frac{\gamma}{\sigma^2}\right)^+$ and $\frac{(e^{-z}-1)^2-(z\wedge 1)^2}{2\sigma^2 t^2} \leq 0$. Moreover,

$$\begin{aligned} \frac{\gamma z}{\sigma^2 t} - \frac{z^2 - (z\wedge 1)^2}{2\sigma^2 t^2} &= \frac{1}{\sigma^2}\left(\gamma\frac{z}{t} - \frac{1}{2}\left(\frac{z}{t}\right)^2 + \frac{(\frac{z}{t}\wedge\frac{1}{t})^2}{2}\right) \\ &\leq \begin{cases} \frac{\gamma}{\sigma^2 t} & \text{if } \gamma t \leq 1 \\ \frac{\gamma^2}{\sigma^2} & \text{otherwise} \end{cases} \end{aligned}$$

   Consequently, $\phi^+(t,z) \leq \frac{\gamma^2}{\sigma^2} + \frac{\gamma}{\sigma^2 t} + \frac{1}{t}\left(\frac{1}{2} - \frac{\gamma}{\sigma^2}\right)^+$.
2. $\underline{\gamma \leq 0}$
   Now we rewrite ϕ in the following form

$$\phi(t,z) = \frac{e^{-z} - 1 + z - \frac{z^2}{2}}{\sigma^2 t^2} + \gamma\frac{e^{-z} - 1 + z}{\sigma^2 t} + \frac{(e^{-z}-1)^2 - z^2}{2\sigma^2 t^2} + \frac{1 - e^{-z}}{2t}$$

   It is then easy to show that $\phi^+(t,z) \leq \frac{1}{2t}$.

Note that $\frac{1}{2t} \leq \frac{\gamma^2}{\sigma^2} + \frac{\gamma}{\sigma^2 t} + \frac{1}{t}\left(\frac{1}{2} - \frac{\gamma}{\sigma^2}\right)^+$. Hence, gathering the two cases yields the first part of the lemma.

Let now $z \in [c, 0]$ for a given negative constant c. We rewrite ϕ in the following form

$$\phi(t,z) = \underbrace{\frac{e^{-z}-1+z}{\sigma^2 t^2}(1+\gamma^+ t) + \frac{(e^{-z}-1)^2}{2\sigma^2 t^2} - \frac{z^2}{\sigma^2 t^2}}_{\geq 0 \text{ for } z<0} + \underbrace{\frac{1-e^{-z}}{2t} - \gamma^- \frac{e^{-z}-1+z}{\sigma^2 t}}_{\leq 0 \text{ for } z<0}.$$

Since $\partial_z \left[\frac{e^{-z}-1+z}{\sigma^2 t^2}(1+\gamma^+ t) + \frac{(e^{-z}-1)^2}{2\sigma^2 t^2} - \frac{z^2}{\sigma^2 t^2} \right] = \frac{1-e^{-2z}-2z+t\gamma^+(1-e^{-z})}{t^2\sigma^2}$ is negative for all $z < 0$, one has that

$$\sup_{z\in[c,0]} \phi^+(t,z) \leq \frac{e^{-c}-1+c}{\sigma^2 t^2}(1+\gamma^+ t) + \frac{(e^{-c}-1)^2}{2\sigma^2 t^2} - \frac{c^2}{\sigma^2 t^2}.$$

□

This lemma suggests to apply the exact algorithm on $[\epsilon, T]$ for a fixed positive threshold ϵ. It remains to handle the time interval $[0, \epsilon[$. Thanks to the following lemma, we that $\phi^+(t, Z_t)$ can be approximately bounded from above for small t, almost surely, by a function of t. The idea is then to extend the exact simulation algorithm by simulating an inhomogeneous Poisson process. Of course, this hybrid method is no longer exact since the positive threshold for which the upper bound holds is random.

Lemma 17 — *For all $\eta > 0$, there exists a random neighborhood of $t = 0$ such that*

$$\phi^+(t, Z_t) \leq \left(\frac{2c^3}{3\sigma^2} + \frac{c}{2} \right) t^{-\frac{1}{2}-\eta} \tag{1.24}$$

where $c = \max(\frac{\sigma}{3^{\frac{1}{2}-\frac{\eta}{3}}}, \frac{\gamma}{2})$.

Proof : We rewrite (1.19) this way

$$\phi(t,z) = \left(\frac{1-e^{-z}}{2} + \gamma \frac{e^{-z}-1+z}{\sigma^2} \right) \frac{1}{t} - \left(\frac{1-z+\frac{z^2}{2}-e^{-z}-\frac{1}{2}(e^{-z}-1+z)(e^{-z}-1-z)}{\sigma^2} \right) \frac{1}{t^2}$$

and make the following Taylor expansions

$$\frac{1-z+\frac{z^2}{2}-e^{-z}-\frac{1}{2}(e^{-z}-1+z)(e^{-z}-1-z)}{\sigma^2} = \frac{2}{3\sigma^2} z^3 + \mathcal{O}(z^4)$$

$$\text{and } \frac{1-e^{-z}}{2} + \gamma \frac{e^{-z}-1+z}{\sigma^2} = \frac{1}{2} z + \mathcal{O}(z^2).$$

On the other hand, we have seen in the proof of lemma 11 that there exists a random neighborhood of zero such that $Z_t \leq c t^{\frac{1}{2}-\eta}$ where $c = \max(\frac{\sigma}{3^{\frac{1}{2}-\frac{\eta}{3}}}, \frac{\gamma}{2})$. We conclude that, in a random neighborhood of zero,

$$\phi^+(t, Z_t) \leq \left(\frac{2c^3}{3\sigma^2} + \frac{c}{2} \right) t^{-\frac{1}{2}-\eta}.$$

□

Numerical computation

For numerical computation, we are going to use the following set of parameters : $S_0 = 100$, $K = 100$, $\sigma = 0.2$, $r = 0.1$, $\delta = 0$ and $T = 1$. To fix the ideas, let us consider a call option. The price C_0 writes as follows

$$C_0 = \mathbb{E}\left(e^{A(T,Z_T)-rT}\left(S_0 e^{Z_T} - K\right)^+ \left(e^{c_p}\prod_{i=1}^{N}\frac{\phi^-(U_i, Z_{U_i})}{c_p}\right) e^{-\int_0^T \phi^+(t,Z_t)dt}\right).$$

where $N \sim \mathcal{P}(c_p)$ and $(U_i)_{i\geq 1}$ is an independent sequence of independent random variables uniformly distributed in $[0, T]$. The parameter $c_p > 0$ is set to one in the following. We give a description of the hybrid method we implement :

Algorithm 2

On the time interval $I_j := [\frac{T}{2^{j+1}}, \frac{T}{2^j}]$,

1. *Simulate $Z_{\frac{T}{2^{j+1}}}, Z_{\frac{T}{2^j}}$ and a lower bound m_j for the minimum of $(Z_t)_{t\in I_j}$ (use the fact that $Z_t = \frac{\sigma}{t}B_{\frac{t^3}{3}} + \frac{\gamma}{2}t$ where $(B_t)_{t\geq 0}$ is a standard Brownian motion).*
2. *Find $M^j > 0$ such that $\forall t \in I_j, \phi(t, Z_t) \leq M^j$ (use Lemma 16).*
3. *Simulate an homogeneous spatial Poisson process on the rectangle $I_j \times [0, M^j]$ and accept (respectively reject) the trajectory simulated if the number of points falling below the graph $(\phi^+(t, Z_t))_{t\in I_j}$ is equal to (respectively different from) zero.*

Carry on this acceptance rejection algorithm until reaching a time interval I_J for a chosen $J \in \mathbb{N}^*$. On the remaining time interval $[0, \frac{T}{2^{J+1}}]$, use the same acceptance/rejection algorithm but with an inhomogeneous spatial Poisson process this time (use Lemma 17).

In table 1.3, we give the price obtained by our method for different values of the positive threshold $\epsilon = \frac{T}{2^{J+1}}$. The number M of Monte Carlo simulations is equal to 10^5 and the true price is equal to 7.042 (computed using a Monte Carlo method with a trapezoidal scheme and a Kemna-Vorst control variate technique).

	Price	CPU
$\epsilon = \frac{T}{2^2}$	6.9394	7s
$\epsilon = \frac{T}{2^4}$	6.9590	10s
$\epsilon = \frac{T}{2^6}$	6.9703	13s
$\epsilon = \frac{T}{2^8}$	6.9952	17s
$\epsilon = \frac{T}{2^{10}}$	7.0423	21s

Table 1.3: Price of the Asian call using the hybrid-pseudo exact method.

Clearly, the method is not yet competitive regarding computation time. Nevertheless, unlike the usual discretization methods, it is not prone to discretization errors.

1.3 Conclusion

In this article, we have applied two original Monte Carlo methods for pricing Asian like options which have the following pay-off : $(\alpha S_T + \beta \int_0^T S_t dt - K)_+$. In the case $\alpha \neq 0$, we applied both

the algorithm of Beskos *et al.* [13] and a method based on the unbiased estimator of Wagner [126] and more recently the Poisson estimator of Beskos *et al.* [14] and the generalized Poisson estimator of Fearnhead *et al.* [38]. The numerical results show that the latter performs the best. The more interesting case $\alpha = 0$, which corresponds to usual continuously monitored Asian options, can not be treated using neither the exact algorithm, nor the method of exact computation of expectation but we investigate an hybrid pseudo-exact method which combines the two techniques. More generally, this hybrid method is an extension of the two exact methods and can be applied in other situations.

From a practical point of view, the main contribution of these techniques is to allow Monte Carlo pricing without resorting to discretization schemes. Hence, we are no longer prone to the discretization bias that we encounter in standard Monte Carlo methods for pricing Asian like options. Even though these exact methods are time consuming, they provide a good and reliable benchmark.

1.4 Appendix

1.4.1 The practical choice of p and q in the U.E method

The best choice for the probability law p of N and the common density q of the variables $(V_i)_{i\geq 1}$ is obviously the one for which the variance of the simulation is minimum. In a very general setting, it is difficult to tackle this issue. In order to have a first idea, we are going to restrict ourselves to the computation of $\mathbb{E}\left(\frac{1}{p(N)\,N!}\prod_{i=1}^{N}\frac{g(V_i)}{q(V_i)}\right)$ where $g : [0,T] \to \mathbb{R}$.

Lemma 18 — *When g is a measurable function on $[0,T]$ such that $0 < \int_0^T |g(t)|dt < +\infty$, the variance of $\frac{1}{p(N)\,N!}\prod_{i=1}^{N}\frac{g(V_i)}{q(V_i)}$ is minimal for*

$$q_{opt}(t) = \frac{|g(t)|}{\int_0^T |g(t)|dt}\,\mathbb{1}_{[0,T]}(t) \text{ and } p_{opt}(n) = \frac{\left(\int_0^T |g(t)|dt\right)^n}{n!}\exp\left(-\int_0^T |g(t)|dt\right).$$

Proof : Minimizing the variance in (1.7) comes down to minimizing the expectation of the square of $\frac{1}{p(N)\,N!}\prod_{i=1}^{N}\frac{g(V_i)}{q(V_i)}$.

Set

$$F(p,q) = \mathbb{E}\left(\frac{1}{(p(N)\,N!)^2}\prod_{i=1}^{N}\frac{g^2(V_i)}{q^2(V_i)}\right) = \sum_{n=0}^{+\infty}\frac{\left(\int_0^T \frac{g^2(t)}{q(t)}dt\right)^n}{p(n)\,(n!)^2}.$$

Using Cauchy-Schwartz inequality we obtain a lower bound for $F(p,q)$

$$\begin{aligned}
F(p,q) = \sum_{n=0}^{+\infty} \left(\frac{\left(\int_0^T \frac{g^2(t)}{q(t)} dt\right)^{\frac{n}{2}}}{p(n)\, n!} \right)^2 p(n) &\geq \left(\sum_{n=0}^{+\infty} \frac{\left(\int_0^T \frac{g^2(t)}{q(t)} dt\right)^{\frac{n}{2}}}{n!} \right)^2 \\
&= \left(\sum_{n=0}^{+\infty} \frac{\left(\int_0^T \left(\frac{g(t)}{q(t)}\right)^2 q(t) dt\right)^{\frac{n}{2}}}{n!} \right)^2 \\
&\geq \left(\sum_{n=0}^{+\infty} \frac{\left(\int_0^T |g(t)| dt\right)^{n}}{n!} \right)^2 \\
&= \exp\left(2 \int_0^T |g(t)| dt\right).
\end{aligned}$$

We easily check that this lower bound is attained for q_{opt} and p_{opt}.

□

The optimal probability distribution p_{opt} is the Poisson law with parameter $\int_0^T |g(t)| dt$. This justifies our use of a Poisson distribution for p.

1.4.2 Simulation from the distribution h given by (1.13)

Recall that

$$h(u) = C \exp\left(A(u) - \frac{(u - X_0)^2}{2T}\right) = C \exp\left(\frac{\gamma}{\sigma} u + \frac{\beta S_0}{\sigma^2}(1 - e^{-\sigma u}) - \frac{(u - X_0)^2}{2T}\right)$$

where C is a normalizing constant.

The expansion of the exponential $e^{-\sigma u}$ at the first order yields

$$h(u) \approx C \exp\left(\frac{\gamma}{\sigma} u + \frac{\beta S_0}{\sigma} u - \frac{(u - X_0)^2}{2T}\right) = C \exp\left(-\frac{(u - (X_0 + \frac{T(\gamma + \beta S_0)}{\sigma}))^2}{2T}\right).$$

This suggests to do rejection sampling using the normal distribution with mean $X_0 + \frac{T(\gamma+\beta S_0)}{\sigma}$ and variance T as prior. Unfortunately, for a standard set of parameters, this method gives bad results. Even a second order expansion of $e^{-\sigma u}$ which also modifies the variance does not work.

In order to get round this problem, we evaluate the mode u^* of h. We have

$$h'(u^*) = C\left(\frac{\gamma}{\sigma} + \frac{\beta S_0}{\sigma} e^{-\sigma u^*} - \frac{u^* - X_0}{T}\right) \exp\left(\frac{\gamma}{\sigma} u^* + \frac{\beta S_0}{\sigma^2}(1 - e^{-\sigma u^*}) - \frac{(u^* - X_0)^2}{2T}\right).$$

So, $h'(u^*) = 0$ if and only if

$$\frac{\gamma}{\sigma} + \frac{\beta S_0}{\sigma} e^{-\sigma u^*} - \frac{u^* - X_0}{T} = 0$$

which writes

$$\sigma(u^* - X_0 - \frac{\gamma}{\sigma}T)e^{\sigma(u^* - X_0 - \frac{\gamma}{\sigma}T)} = T\beta S_0 e^{-\sigma X_0 - \gamma T}.$$

The function $x \mapsto xe^x$ is continuous and increasing on $[0, +\infty[$ and so is its inverse which we denote by W. Since $T\beta S_0 e^{-\sigma X_0 - \gamma T} \geq 0$, we deduce that h is unimodal and that its mode satisfies

$$u^* = \frac{\gamma T + W\left(\beta S_0 T e^{-\gamma T - \sigma X_0}\right) + \sigma X_0}{\sigma}.$$

The function W is the well-known Lambert function, also called the Omega function. It is uniquely valued on $[0, +\infty[$ and there are robust and fast numerical methods based on series expansion for approximating this function (see for example Corless *et al.* [24]).

Numerical tests showed that performing rejection sampling using a Gaussian distribution with variance T and mean u^* instead of $X_0 + \frac{T(\gamma+\beta S_0)}{\sigma}$ gives plain satisfaction. In table 1.4, we see that for arbitrary choice of the parameter $\frac{\alpha}{\alpha+\beta}$, the acceptance rate of the algorithm is always high (of order 70%) and that the computation time is low.

$\frac{\alpha}{\alpha+\beta}$	Nb of simulations	Acceptance rate	Computation time
0.2	10^6	61%	3s
0.5		68%	3s
0.8		80%	2s

Table 1.4: Acceptance rate of the rejection algorithm of simulating from the distribution h in (1.13) with $S_0 = 100, \sigma = 0.3, T = 2$ and $r = 0.1$.

Chapitre 2

Schémas de discrétisation pour modèles à volatilité stochastique

Ce chapitre est un article écrit avec mon directeur de thèse Benjamin Jourdain. Il a été soumis pour publication.

Abstract. In usual stochastic volatility models, the process driving the volatility of the asset price evolves according to an autonomous one-dimensional stochastic differential equation. We assume that the coefficients of this equation are smooth. Using Itô's formula, we get rid, in the asset price dynamics, of the stochastic integral with respect to the Brownian motion driving this SDE. Taking advantage of this structure, we propose

- a scheme, based on the Milstein discretization of this SDE, with order one of weak trajectorial convergence for the asset price,
- a scheme, based on the Ninomiya-Victoir discretization of this SDE, with order two of weak convergence for the asset price.

We also propose a specific scheme with improved convergence properties when the volatility of the asset price is driven by an Ornstein-Uhlenbeck process. We confirm the theoretical rates of convergence by numerical experiments and show that our schemes are well adapted to the multilevel Monte Carlo method introduced by Giles [48, 47].

Introduction

There exists an extensive literature on numerical integration schemes for stochastic differential equations. To start with, we mention, among many others, the work of Talay and Tubaro [117] who first established an expansion of the weak error of the Euler scheme for polynomially growing functions allowing for the use of Romberg extrapolation. Bally and Taley [7] extended this result to bounded measurable functions and Guyon [52] extended it to tempered stable distributions. More recently, many discretization schemes of higher weak convergence order have appeared in the literature. Among others, we cite the work of Kusuoka [76, 77], the Ninomiya and Victoir [98] scheme which we will use hereafter, the Ninomiya and Ninomiya [97] scheme and the scheme based on cubature on Wiener spaces of Lyons and Victoir [87].
Concerning strong approximation, the Milstein scheme has order one of strong convergence. Unfortunately, it involves the simulation of iterated Brownian integrals unless a restrictive commutativity condition is satisfied. Under ellipticity, Cruzeiro *et al.* [27] have recently proposed a discretization scheme which gets rid of these iterated integrals and has nice strong convergence properties. More precisely, for each number of time steps, there exists a Brownian motion different from the one giving the Brownian increments involved in the scheme such that the strong error between the scheme and the stochastic differential equation driven by this new Brownian motion is of order one. We call such a property weak trajectorial convergence of order one. Weak trajectorial error estimation is exactly what is needed to control the discretization bias for the computation of path dependent option prices.

Stochastic volatility models, which have now become a standard of the market, are an eloquent example of the use of stochastic differential equations in finance. In our study, we will consider the following specification of a stochastic volatility model for an asset $(S_t)_{t\in[0,T]}$:

$$\left\{\begin{array}{rcl} dS_t &=& rS_t dt + f(Y_t)S_t\left(\rho dW_t + \sqrt{1-\rho^2}dB_t\right); \quad S_0 = s_0 > 0 \\ dY_t &=& b(Y_t)dt + \sigma(Y_t)dW_t; \quad Y_0 = y_0 \end{array}\right. \tag{2.1}$$

where r the instantaneous interest rate, $(B_t)_{t\in[0,T]}$ and $(W_t)_{t\in[0,T]}$ are independent standard one-dimensional Brownian motions, $\rho \in [-1,1]$ is the correlation between the Brownian motions respectively driving the asset price and the process $(Y_t)_{t\in[0,T]}$ which solves a one-dimensional autonomous stochastic differential equation. The volatility process is $(f(Y_t))_{t\in[0,T]}$ where the transformation function f is usually taken positive and strictly monotonic in order to ensure that the effective correlation between the stock price and the volatility keeps the same sign (the function σ usually takes nonnegative values). This specification nests almost all the known stochastic volatility models:

- Hull&White model [57] ($\rho = 0$) and Wiggins [131] ($\rho \neq 0$)

$$\left\{\begin{array}{rcl} dS_t &=& rS_t dt + \sqrt{Y_t}S_t\left(\rho dW_t + \sqrt{1-\rho^2}dB_t\right) \\ dY_t &=& \mu Y_t dt + \zeta Y_t dW_t \end{array}\right.$$

  which can be expressed as (2.1) with $f(y) = \sqrt{y}$, $b(y) = \mu y$ and $\sigma(y) = \zeta y$. Note that it can also be seen as (2.1) with $f(y) = e^y$, $b(y) = \frac{\mu}{2} - \frac{\zeta^2}{4}$ and $\sigma(y) = \frac{\zeta}{2}$.
- Scott's model [112] which generalizes Hull&White model

$$\left\{\begin{array}{rcl} dS_t &=& rS_t dt + e^{Y_t}S_t\left(\rho dW_t + \sqrt{1-\rho^2}dB_t\right) \\ dY_t &=& \kappa(\theta - Y_t)dt + \nu dW_t \end{array}\right. \tag{2.2}$$

$$\Rightarrow f(y) = e^y,\ b(y) = \kappa(\theta - y) \text{ and } \sigma(y) = \nu.$$

– Stein&Stein model [115]

$$\left\{\begin{array}{rcl} dS_t &=& rS_t dt + Y_t S_t \left(\rho dW_t + \sqrt{1-\rho^2} dB_t\right) \\ dY_t &=& \kappa(\theta - Y_t)dt + \nu dW_t \end{array}\right.$$
$$\Rightarrow f(y) = y,\ b(y) = \kappa(\theta - y) \text{ and } \sigma(y) = \nu.$$

– Quadratic Gaussian model

$$\left\{\begin{array}{rcl} dS_t &=& rS_t dt + Y_t^2 S_t \left(\rho dW_t + \sqrt{1-\rho^2} dB_t\right) \\ dY_t &=& \kappa(\theta - Y_t)dt + \nu dW_t \end{array}\right.$$
$$\Rightarrow f(y) = y^2,\ b(y) = \kappa(\theta - y) \text{ and } \sigma(y) = \nu.$$

– Heston model [55]

$$\left\{\begin{array}{rcl} dS_t &=& rS_t dt + \sqrt{Y_t} S_t \left(\rho dW_t + \sqrt{1-\rho^2} dB_t\right) \\ dY_t &=& \kappa(\theta - Y_t)dt + \nu\sqrt{Y_t} dW_t \end{array}\right.$$
$$\Rightarrow f(y) = \sqrt{y},\ b(y) = \kappa(\theta - y) \text{ and } \sigma(y) = \nu\sqrt{y}.$$

In all but the last example, the volatility of the asset is driven by an Ornstein Uhlenbeck process.

The development of specific discretization schemes for stochastic volatility models has only received little attention. We mention nevertheless the work of Kahl and Jäckel [61] who discussed various numerical integration methods and proposed a simple scheme with order 1/2 of strong convergence like the standard Euler scheme but with a smaller multiplicative constant. Also the numerical integration of the CIR process and of the Heston model received a particular attention because of the inadequacy of the Euler scheme due to the fact that both f and σ are equal to the square root function (see for example Deelstra and Delbaen [29], Alfonsi [1], Kahl and Schurz [62], Andersen [3], Berkaoui *et al.* [12], Ninomiya and Victoir [98], Lord *et al.* [86], Alfonsi [2]). An exact simulation technique for the Heston model was also proposed by Broadie and Kaya [20].

In the present paper, we assume in return that the functions f, σ and b are smooth and do not deal with the Heston model. Our aim is to take advantage of the structure of (2.1) to construct and analyse simple and robust ad'hoc discretization schemes which have nice convergence properties.

For a start, we make a logarithmic change of variables for the asset : the two-dimensional process $(X_t := \log(S_t), Y_t)_{t\in[0,T]}$ solves the following SDE

$$\left\{\begin{array}{rcl} dX_t &=& (r - \frac{1}{2}f^2(Y_t))dt + f(Y_t)\left(\rho dW_t + \sqrt{1-\rho^2} dB_t\right); \quad X_0 = \log(s_0). \\ dY_t &=& b(Y_t)dt + \sigma(Y_t)dW_t; \quad Y_0 = y_0 \end{array}\right. \tag{2.3}$$

Our main idea is to get rid in the first equality of the stochastic integral involving the common Brownian motion $(W_t)_{t\in[0,T]}$. In all what follows, we assume that

$(\mathcal{A})$ f and σ are $\mathcal{C}^1$ functions and $\sigma > 0$.

One can then define the primitive $F(y) = \int_0^y \frac{f}{\sigma}(z)dz$ and apply Itô's formula to get

$$dF(Y_t) = \frac{f}{\sigma}(Y_t)dY_t + \frac{1}{2}(\sigma f' - f\sigma')(Y_t)dt.$$

Therefore $(X_t, Y_t)_{t\in[0,T]}$ solves

$$\left\{ \begin{array}{rcl} dX_t &=& \rho dF(Y_t) + h(Y_t)dt + \sqrt{1-\rho^2} f(Y_t)dB_t \\ dY_t &=& b(Y_t)dt + \sigma(Y_t)dW_t \end{array} \right. \tag{2.4}$$

where $h : y \mapsto r - \frac{1}{2}f^2(y) - \rho(\frac{b}{\sigma}f + \frac{1}{2}(\sigma f' - f\sigma'))(y)$. We discretize the autonomous SDE satisfied by Y using a scheme with high order of strong or weak convergence depending on wether one is interested in path-dependent or vanilla options. Then, in the dynamics of X, we only need to discretize the standard integral $\int_0^T h(Y_s)ds$ and the stochastic integral $\int_0^T f(Y_t)dB_t$ where $(Y_t)_{t\in[0,T]}$ and $(B_t)_{t\in[0,T]}$.

We recall that usual weak convergence is the right notion to analyse the discretization bias for plain vanilla options whereas weak trajectorial convergence permits to deal with path-dependent options. The first section of the paper is devoted to path-dependent options. Combining the Milstein discretization of the one-dimensional SDE satisfied by $(Y_t)_{t\in[0,T]}$ with an appropriate discretization of the integral $\int_0^T f(Y_t)dB_t$ based on the independence of $(Y_t)_{t\in[0,T]}$ and $(B_t)_{t\in[0,T]}$, we obtain a scheme with order one of weak trajectorial convergence. In the second section, using the Ninomiya-Victoir discretization of the SDE satisfied by $(Y_t)_{t\in[0,T]}$, we construct a scheme with order two of weak convergence. Since the SDE satisfied by Y is one-dimensional, the Ninomiya-Victoir scheme only involves two one-dimensional ODEs whose solutions are available in closed form. The last section is devoted to numerical experiments which confirm the theoretical rates of convergence. We also show that our schemes are well adapted to the multilevel Monte Carlo method introduced by Giles [48, 47].

Notations

We will consider, for a number of time steps $N \geq 1$, the uniform subdivision $\prod_N = \{0 = t_0 < t_1 < \cdots < t_N = T\}$ of $[0,T]$ with the discretization step $\delta_N = \frac{T}{N}$.
We denote by $\underline{\psi}$ the greatest lower bound of the function $\psi : y \mapsto f^2(y)$ and by $\overline{\psi}$ its lowest upper bound. We also introduce the following notation :

$$\widehat{\psi}(y) = \left\{ \begin{array}{ll} \frac{3}{2}f^2(y) & \text{if } \overline{\psi} = \infty \\ \overline{\psi} & \text{otherwise} \end{array} \right.$$

2.1 An efficient scheme for path dependent options pricing

Building a first order strong convergence scheme for a two dimensional SDE is not an obvious task. Even the ad'hoc schemes provided by Kahl and Jäckel [61] exhibit a strong convergence of order $\frac{1}{2}$.

Actually, the natural candidate for this purpose is the Milstein scheme. Unfortunately, the commutativity condition which permits to implement it amounts to $\sigma f' = 0$ in our setting. This condition is typically true when either f is constant or $\sigma = 0$. Both cases are of no practical interest since they lead to a deterministic volatility.

However, since the inherent Brownian motion is not essential for applications in finance, the usual strong convergence criterion is not adapted for estimating the error of a scheme in pricing a path dependent option. What is more relevant is the approximation in law of the whole trajectory

of the process considered for instance by Cruzeiro *et al.* [27]. Using an ingenious rotation of the Brownian motion, these authors have constructed a discretization scheme allowing for a weak convergence on the whole trajectory of order one which avoids the simulation of the iterated stochastic integrals.

For the SDE (2.3), the discretization scheme of Cruzeiro, Malliavin and Thalmaier writes as

$$\begin{aligned} X_{t_{k+1}}^{CMT} &= X_{t_k}^{CMT} + \left(r - \frac{f^2(Y_{t_k}^{CMT})}{2}\right)\delta_N + \rho f(Y_{t_k}^{CMT})\Delta W_{k+1} + \tfrac{\rho}{2}\sigma f'(Y_{t_k}^{CMT})\Delta W_{k+1}^2 \\ &\quad + \sqrt{1-\rho^2}\sigma f'(Y_{t_k}^{CMT})\Delta W_{k+1}\Delta B_{k+1} + \sqrt{1-\rho^2} f(Y_{t_k}^{CMT})\Delta B_{k+1} - \tfrac{\rho}{2}\sigma f'(Y_{t_k}^{CMT})\Delta B_{k+1}^2 \\ Y_{t_{k+1}}^{CMT} &= Y_{t_k}^{CMT} + \left(b(Y_{t_k}^{CMT}) + \tfrac{1}{2}(\tfrac{\sigma^2 f'}{f} - \sigma\sigma')(Y_{t_k}^{CMT})\right)\delta_N + \sigma(Y_{t_k}^{CMT})\Delta W_{k+1} \\ &\quad + \tfrac{1}{2}\sigma\sigma'(Y_{t_k}^{CMT})\Delta W_{k+1}^2 - \tfrac{\sigma^2 f'}{2f}\Delta B_{k+1}^2 \end{aligned} \tag{2.5}$$

where $\Delta W_{t_{k+1}} = W_{t_{k+1}} - W_{t_k}$ and $\Delta B_{k+1} = B_{t_{k+1}} - B_{t_k}$ correspond to the Brownian increments.

We set out to construct a much simpler scheme having the same order of weak trajectorial convergence by taking advantage of the particular structure of the SDE defining stochastic volatility models. We first begin with the general case of any process $(Y_t)_{t\in[0,T]}$ driving the volatility and then consider the case of an Ornstein-Uhlenbeck process where we obtain more precise results.

2.1.1 General case

A discretization scheme will naturally involve the Brownian increments. Thanks to the independence between $(Y_t)_{t\in[0,T]}$ and $(B_t)_{t\in[0,T]}$, we can construct a vector $(\widetilde{X}_{t_0},\dots,\widetilde{X}_{t_N})$ using only $(\Delta B_1,\dots,\Delta B_N)$ and $(Y_t)_{t\in[0,T]}$, which has exactly the same law as $(X_{t_0},\dots,X_{t_N})$:

Lemma 19 — *$\forall 0 \le l < N$, let $v_l = \frac{1}{\delta_N}\int_{t_l}^{t_{l+1}}\psi(Y_s)ds$. The vector $(\widetilde{X}_{t_0},\dots,\widetilde{X}_{t_N})$ defined by*

$$\begin{aligned} &\widetilde{X}_{t_0} = X_{t_0} \\ &\forall 1 \le k \le N, \widetilde{X}_{t_k} = \widetilde{X}_{t_0} + \rho(F(Y_{t_k}) - F(Y_{t_0})) + \int_0^{t_k} h(Y_s)ds + \sqrt{1-\rho^2}\sum_{l=0}^{k-1}\sqrt{v_l}\,\Delta B_{l+1} \end{aligned}$$

has the same law as $(X_{t_0},\dots,X_{t_N})$.

Proof : The proof is elementary. Conditionally on Y, the two vectors are Gaussian vectors with the same mean and covariance matrix. □

In order to approximate $(\widetilde{X}_{t_k})_{0\le k\le N}$, one needs to discretize v_k for $k \in \{0,\dots,N-1\}$. If $(v_k^N)_{0\le k\le N-1}$ is an approximation of $(v_k)_{0\le k\le N-1}$, then by Doob's inequality

$$\begin{aligned} \mathbb{E}\left[\sup_{0\le k\le N-1}\left(\sum_{l=0}^{k}\left(\sqrt{v_l} - \sqrt{v_l^N}\right)\Delta B_{l+1}\right)^2\right] &\le 4\delta_N\sum_{k=0}^{N-1}\mathbb{E}\left[\left(\sqrt{v_l} - \sqrt{v_l^N}\right)^2\right] \\ &\le \frac{1}{\underline{\psi}}\delta_N\sum_{k=0}^{N-1}\mathbb{E}\left[\left(v_l - v_l^N\right)^2\right] \end{aligned}$$

as soon as $\underline{\psi} = \inf_x \psi(x)$ is assumed to be positive and, $\forall 0 \leq k \leq N-1, v_k^N$ is greater than $\underline{\psi}$. Consequently, to obtain a scheme with order one of strong convergence for $(\widetilde{X}_{t_k})_{0\leq k\leq N}$, one needs that $\forall 0 \leq k \leq N-1, \mathbb{E}\left[\left(v_k - v_k^N\right)^2\right] = \mathcal{O}\left(\frac{1}{N^2}\right)$. According to the treatment of the term $\overline{I}_2^j$ defined by (2.9) in the proof of the Theorem 20 below, one has $\forall 0 \leq k \leq N-1$,

$$\mathbb{E}\left[\left(v_k - \left(\psi(Y_{t_k}) + \frac{\sigma\psi'(Y_{t_k})}{\delta_N}\int_{t_k}^{t_{k+1}}(W_s - W_{t_k})ds\right)\right)^2\right] = \mathcal{O}\left(\frac{1}{N^2}\right). \tag{2.6}$$

This equality still holds true when replacing Y by a scheme with order one of strong convergence in the term with sign minus of the left hand side. Better still, $\left(F(Y_{t_k}) + \int_0^{t_k} h(Y_s)ds\right)_{0\leq k\leq N}$ is approximated with strong order one when replacing Y by such a scheme and using a rectangular discretization for the integral in time.

For all these reasons, we choose the Milstein scheme for Y :

$$\forall 0 \leq k \leq N-1,\ \widetilde{Y}_{t_{k+1}}^N = \widetilde{Y}_{t_k}^N + b(\widetilde{Y}_{t_k}^N)\delta_N + \sigma(\widetilde{Y}_{t_k}^N)\Delta W_{k+1} + \frac{1}{2}\sigma\sigma'(\widetilde{Y}_{t_k}^N)\left(\Delta W_{k+1}^2 - \delta_N\right); \quad \widetilde{Y}_{t_0}^N = y_0.$$

and we write our scheme as follows

WeakTraj_1 scheme

$$\begin{aligned}\widetilde{X}_{t_{k+1}}^N &= \widetilde{X}_{t_k}^N + \rho\left(F(\widetilde{Y}_{t_{k+1}}^N) - F(\widetilde{Y}_{t_k}^N)\right) + \delta_N h(\widetilde{Y}_{t_k}^N)\\ &+\sqrt{1-\rho^2}\sqrt{\left(\psi(\widetilde{Y}_{t_k}^N) + \frac{\sigma\psi'(\widetilde{Y}_{t_k}^N)}{\delta_N}\int_{t_k}^{t_{k+1}}(W_s - W_{t_k})ds\right) \vee \underline{\psi}}\ \Delta B_{k+1}\end{aligned} \tag{2.7}$$

Note that in order to implement this scheme, one needs to simulate both the Brownian increment ΔW_{k+1} and the random variable $\int_{t_k}^{t_{k+1}}(W_s - W_{t_k})ds$. This is straightforward as one can easily check that

$$\begin{pmatrix} \Delta W_{k+1} \\ \displaystyle\int_{t_k}^{t_{k+1}}(W_s - W_{t_k})ds \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \delta_N & \delta_N^2/2 \\ \delta_N^2/2 & \delta_N^3/3 \end{pmatrix}\right)$$

We can now state our first main result :

Theorem 20 — *Under the assumptions of Lemma 21 and if*

($\mathcal{H}1$) f *and* σ *are* $\mathcal{C}^3$ *functions,* $\frac{f}{\sigma}$ *and* ff' *are bounded*

($\mathcal{H}2$) $\underline{\psi} > 0$

($\mathcal{H}3$) *there exists a constant* K_1 *such that,* $\forall (x,y) \in \mathbb{R}^2$,

$$\begin{aligned}&\left|(bh' + \frac{\sigma^2}{2}h'')(y)\right| \leq K_1(1+|y|)\\ &\left|\sigma h'(y)\right| \leq K_1(1+|y|)\\ &\left|h(y) - h(x)\right| \leq K_1|y-x|\end{aligned}$$

($\mathcal{H}4$) there exists a constant K_2 such that, $\forall (x,y) \in \mathbb{R}^2$,

$$\left|(b\psi' + \frac{\sigma^2}{2}\psi'')(y)\right| \leq K_2(1+|y|)$$
$$\left|\sigma\psi'(y) - \sigma\psi'(x)\right| \leq K_2|y-x|$$

then the WeakTraj_1 scheme has order one of weak trajectorial convergence. More precisely, for each $p \geq 1$, there exists a constant C independent of the number of time steps N such that

$$\mathbb{E}\left[\max_{0\leq k\leq N}\left\|\left(\widetilde{X}_{t_k}, Y_{t_k}\right) - \left(\widetilde{X}^N_{t_k}, \widetilde{Y}^N_{t_k}\right)\right\|^{2p}\right] \leq \frac{C}{N^{2p}}.$$

The proof of the theorem relies on the order one of strong convergence of the Milstein scheme (see Milstein [92] for the particular case $p=1$) :

Lemma 21 — *Suppose that*

($\mathcal{H}5$) b and σ are $\mathcal{C}^2$ functions with bounded first and second derivatives

($\mathcal{H}6$) there exists a positive constant K such that $\forall (x,y) \in \mathbb{R}^2$

$$|\sigma\sigma'(x) - \sigma\sigma'(y)| \leq K|x-y|$$

then, $\forall p \geq 1$, there exists a positive constant C_p independent of N such that

$$\mathbb{E}\left(\max_{0\leq k\leq N}\left|Y_{t_k} - \widetilde{Y}^N_{t_k}\right|^{2p}\right) \leq C_p\delta_N^{2p}.$$

The proof for general p is postponed to the appendix.

Remark 22 — *Before giving the proof of the theorem, we make a few comments on its assumptions. ($\mathcal{H}1$) implies that h and ψ are $\mathcal{C}^2$ functions which was implicitly assumed in ($\mathcal{H}3$) and ($\mathcal{H}4$). The latter assumptions are expressed in a reduced form. One can check that the following conditions on the coefficients of the original SDE are sufficient for them to hold :*

- *f and σ are bounded $\mathcal{C}^4$ functions with bounded derivatives.*
- *b is a bounded $\mathcal{C}^3$ function with bounded derivatives.*
- *$\exists \sigma_0 > 0$ such that $\forall y \in \mathbb{R}$, $\sigma(y) \geq \sigma_0$.*

Proof of the theorem : Throughout the proof, we denote by C a constant which can change from one line to another while always being independent of N. Thanks to Lemma 21, we just have to control the error on $\widetilde{X}$:

$$\begin{aligned}
\mathbb{E}\left[\max_{0\leq k\leq N}|\widetilde{X}_{t_k} - \widetilde{X}^N_{t_k}|^{2p}\right] &= \mathbb{E}\Bigg[\max_{0\leq k\leq N}\Bigg|\rho(F(Y_{t_k}) - F(\widetilde{Y}^N_{t_k})) + \sum_{j=0}^{k-1}\Bigg(\int_{t_j}^{t_{j+1}} h(Y_s)ds - \delta_N h(\widetilde{Y}^N_{t_j}) \\
&\quad + \sqrt{\frac{1-\rho^2}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds}\,\Delta B_{j+1} \\
&\quad - \sqrt{1-\rho^2}\sqrt{\left(\psi(\widetilde{Y}^N_{t_j}) + \frac{\sigma\psi'(\widetilde{Y}^N_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s - W_{t_j})ds\right) \vee \underline{\psi}}\;\Delta B_{j+1}\Bigg|^{2p}\Bigg)\Bigg] \\
&\leq 3^{2p-1}\left(\rho^{2p}I_0 + I_1 + (1-\rho^2)^p I_2\right)
\end{aligned}$$

where

$$I_0 = \mathbb{E}\left[\max_{0\le k\le N}\left|F(Y_{t_k}) - F(\widetilde{Y}^N_{t_k})\right|^{2p}\right]$$

$$I_1 = \mathbb{E}\left[\max_{0\le k\le N}\Big|\sum_{j=0}^{k-1}\left(\int_{t_j}^{t_{j+1}} h(Y_s)ds - \delta_N h(\widetilde{Y}^N_{t_j})\right)\Big|^{2p}\right]$$

and

$$I_2 = \mathbb{E}\left[\max_{0\le k\le N}\Big|\sum_{j=0}^{k-1}\left(\sqrt{\frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds} - \right.\right.$$
$$\left.\left.\sqrt{\left(\psi(\widetilde{Y}^N_{t_j}) + \frac{\sigma\psi'(\widetilde{Y}^N_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s - W_{t_j})ds\right)\vee\underline{\psi}}\right)\Delta B_{j+1}\Big|^{2p}\right].$$

$(\mathcal{H}1)$ yields that F is Lipschitz continuous so using Lemma 21 we show that $I_0 \le \frac{C}{N^{2p}}$. Next, we have that

$$I_1 \le C\left(\mathbb{E}\left[\max_{0\le k\le N}\Big|\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} h(Y_s)ds - \delta_N h(Y_{t_j})\Big|^{2p}\right] + \delta_N^{2p}\mathbb{E}\left[\max_{0\le k\le N}\Big|\sum_{j=0}^{k-1} h(Y_{t_j}) - h(\widetilde{Y}^N_{t_j})\Big|^{2p}\right]\right)$$

On one hand, thanks to assumption $(\mathcal{H}1)$ and Lemma 21,

$$\delta_N^{2p}\mathbb{E}\left[\max_{0\le k\le N}\Big|\sum_{j=0}^{k-1} h(Y_{t_j}) - h(\widetilde{Y}^N_{t_j})\Big|^{2p}\right] \le C\delta_N\sum_{j=0}^{N-1}\mathbb{E}\left[\left|h(Y_{t_j}) - h(\widetilde{Y}^N_{t_j})\right|^{2p}\right] \le \frac{C}{N^{2p}}.$$

On the other hand, using an integration by parts formula,

$$\begin{aligned}
\overline{I}_1 &:= \mathbb{E}\left[\max_{0\le k\le N}\Big|\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} h(Y_s) - h(Y_{t_j})ds\Big|^{2p}\right]\\
&= \mathbb{E}\left[\max_{0\le k\le N}\Big|\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}}(t_{j+1}-s)\left((bh' + \frac{\sigma^2 h''}{2})(Y_s)ds + \sigma h'(Y_s)dW_s\right)\Big|^{2p}\right]\\
&\le 2^{2p-1}\left(\mathbb{E}\left[\max_{0\le k\le N}\Big|\int_0^{t_k}(\tau_s - s)(bh' + \frac{\sigma^2 h''}{2})(Y_s)ds\Big|^{2p}\right] + \mathbb{E}\left[\max_{0\le k\le N}\Big|\int_0^{t_k}(\tau_s - s)\sigma h'(Y_s)dW_s\Big|^{2p}\right]\right)
\end{aligned}$$

where we denoted by τ_s the lowest discretization point greater than s : $\tau_s = \lceil\frac{s}{\delta_N}\rceil\delta_N$. Using Jensen's inequality for the first integral and the Burkholder-Davis-Gundy inequality for the second, we obtain

$$\begin{aligned}
\overline{I}_1 &\le C\left(\mathbb{E}\left[\max_{0\le k\le N} t_k^{2p-1}\int_0^{t_k}(\tau_s - s)^{2p}\left|(bh' + \frac{\sigma^2 h''}{2})(Y_s)\right|^{2p}ds\right]\right.\\
&\qquad\left. + \mathbb{E}\left[\left(\int_0^T(\tau_s - s)^2\left|\sigma h'(Y_s)\right|^2 ds\right)^p\right]\right)\\
&\le \frac{C}{N^{2p}}\int_0^T\mathbb{E}\left[\left|(bh' + \frac{\sigma^2 h''}{2})(Y_s)\right|^{2p} + \left|\sigma h'(Y_s)\right|^{2p}\right]ds.
\end{aligned}$$

Under the assumptions of Lemma 21, $\sup_{0\le t\le T}\mathbb{E}(|Y_s|^{2p})<\infty$ (see Problem 3.15 p. 306 of Karatzas and Shreve [63] for example) so, with the help of assumption $(\mathcal{H}3)$, we conclude that $\overline{I}_1\le\frac{C}{N^{2p}}$ and hence $I_1\le\frac{C}{N^{2p}}$. We now turn to the last term. Using Burkholder-Davis-Gundy inequality, we get

$$\begin{aligned}
I_2 &\le C\delta_N^p\mathbb{E}\left[\left(\sum_{j=0}^{N-1}\left(\sqrt{\frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds}-\sqrt{\left(\psi(\widetilde{Y}_{t_j}^N)+\frac{\sigma\psi'(\widetilde{Y}_{t_j}^N)}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right)\vee\underline{\psi}}\right)^2\right)^p\right]\\
&\le \delta_N\sum_{j=0}^{N-1}\mathbb{E}\left[\left|\sqrt{\frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds}-\sqrt{\left(\psi(\widetilde{Y}_{t_j}^N)+\frac{\sigma\psi'(\widetilde{Y}_{t_j}^N)}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right)\vee\underline{\psi}}\right|^{2p}\right]
\end{aligned}\tag{2.8}$$

Assumption $(\mathcal{H}2)$ yields that the two terms appearing in the square root are bounded from below by $\underline{\psi}>0$ so we have that

$$\begin{aligned}
I_2 &\le C\delta_N\sum_{j=0}^{N-1}\mathbb{E}\left[\left|\frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds-\left(\psi(\widetilde{Y}_{t_j}^N)+\frac{\sigma\psi'(\widetilde{Y}_{t_j}^N)}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right)\vee\underline{\psi}\right|^{2p}\right]\\
&\le CN^{2p-1}\sum_{j=0}^{N-1}\mathbb{E}\left[\left|\int_{t_j}^{t_{j+1}}\psi(Y_s)ds-\left(\psi(\widetilde{Y}_{t_j}^N)\delta_N+\sigma\psi'(\widetilde{Y}_{t_j}^N)\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right)\right|^{2p}\right]\\
&\le CN^{2p-1}\sum_{j=0}^{N-1}\left(\overline{I}_2^j+\widetilde{I}_2^j\right)
\end{aligned}$$

where

$$\overline{I}_2^j=\mathbb{E}\left[\left|\int_{t_j}^{t_{j+1}}\psi(Y_s)ds-\left(\psi(Y_{t_j})\delta_N+\sigma\psi'(Y_{t_j})\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right)\right|^{2p}\right]\tag{2.9}$$

and

$$\widetilde{I}_2^j=\mathbb{E}\left[\left|\delta_N\left(\psi(Y_{t_j})-\psi(\widetilde{Y}_{t_j}^N)\right)+\left(\sigma\psi'(Y_{t_j})-\sigma\psi'(\widetilde{Y}_{t_j}^N)\right)\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right|^{2p}\right]$$

Again, integrating by parts yields that

$$\overline{I}_2^j=\mathbb{E}\left[\left|\int_{t_j}^{t_{j+1}}(t_{j+1}-s)\left((\sigma\psi'(Y_s)-\sigma\psi'(Y_{t_j}))dW_s+((b\psi'+\frac{\sigma^2}{2}\psi'')(Y_s))ds\right)\right|^{2p}\right]$$

We control the stochastic integral term as follows

$$\begin{aligned}
\mathbb{E}\left[\left|\int_{t_j}^{t_{j+1}}(t_{j+1}-s)(\sigma\psi'(Y_s)-\sigma\psi'(Y_{t_j}))dW_s\right|^{2p}\right] &\leq C\delta_N^{p-1}\mathbb{E}\left[\int_{t_j}^{t_{j+1}}(t_{j+1}-s)^{2p}|\sigma\psi'(Y_s)-\sigma\psi'(Y_{t_j})|^{2p}ds\right] \\
&\leq C\delta_N^{3p-1}\int_{t_j}^{t_{j+1}}\mathbb{E}\left[\left|\sigma\psi'(Y_s)-\sigma\psi'(Y_{t_j})\right|^{2p}\right]ds \\
&\leq C\delta_N^{3p-1}\int_{t_j}^{t_{j+1}}\mathbb{E}\left[\left|Y_s-Y_{t_j}\right|^{2p}\right]ds \\
&\leq C\delta_N^{3p-1}\int_{t_j}^{t_{j+1}}|s-t_j|^p\,ds \\
&\leq C\delta_N^{4p}
\end{aligned}$$

The third inequality is due to assumption ($\mathcal{H}4$) and the fourth one is a standard result on the control of the moments of the increments of the solution of a SDE with Lipschitz continuous coefficients (see Problem 3.15 p. 306 of Karatzas and Shreve [63] for example).

We also control the other term thanks to assumption ($\mathcal{H}4$) :

$$\begin{aligned}
\mathbb{E}\left[\left|\int_{t_j}^{t_{j+1}}(t_{j+1}-s)(b\psi'+\frac{\sigma^2}{2}\psi'')(Y_s)ds\right|^{2p}\right] &\leq \delta_N^{2p-1}\mathbb{E}\left[\int_{t_j}^{t_{j+1}}(t_{j+1}-s)^{2p}|(b\psi'+\frac{\sigma^2}{2}\psi'')(Y_s)|^{2p}ds\right] \\
&\leq \delta_N^{4p-1}\int_{t_j}^{t_{j+1}}\mathbb{E}\left[\left|(b\psi'+\frac{\sigma^2}{2}\psi'')(Y_s)\right|^{2p}\right]ds \\
&\leq C\delta_N^{4p}
\end{aligned}$$

Hence, $\overline{I}_2^j \leq \frac{C}{N^{4p}}$. To conclude the proof of the theorem, it remains to show a similar result for $\widetilde{I}_2^j$:

$$\begin{aligned}
\widetilde{I}_2^j &\leq 2^{2p-1}\mathbb{E}\left[\left|\delta_N\left(\psi(Y_{t_j})-\psi(\widetilde{Y}_{t_j}^N)\right)\right|^{2p}+\left|\left(\sigma\psi'(Y_{t_j})-\sigma\psi'(\widetilde{Y}_{t_j}^N)\right)\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right|^{2p}\right] \\
&\leq C\left(\delta_N^{2p}\mathbb{E}\left[\left|Y_{t_j}-\widetilde{Y}_{t_j}^N\right|^{2p}\right]+\frac{\delta_N^{3p}}{3^p}\mathbb{E}\left[\left|Y_{t_j}-\widetilde{Y}_{t_j}^N\right|^{2p}\right]\right) \\
&\leq \frac{C}{N^{4p}}
\end{aligned}$$

The second inequality is due to the fact that ψ is Lipschitz continuous (thanks to assumption ($\mathcal{H}1$)) for the first term and to the independence of $\left(\sigma\psi'(Y_{t_j})-\sigma\psi'(\widetilde{Y}_{t_j}^N)\right)$ and $\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds$ for the second term. □

Remark 23 — *Our scheme exhibits the same convergence properties as the Cruzeiro et al. [27] scheme. Apart from the fact that it involves less terms, it presents the advantage of improving the multilevel Monte Carlo convergence. This method, which is a generalization of the statistical Romberg extrapolation method of Kebaier [65], was introduced by Giles [48, 47].*

Indeed, consider the discretization scheme with time step $\delta_{2N} = \frac{T}{2N}$ *:*

$$\forall 0 \leq k \leq 2N-1,\ \widetilde{X}^{2N}_{\frac{(k+1)T}{2N}} = \widetilde{X}^{2N}_{\frac{kT}{2N}} + \rho\left(F(\widetilde{Y}^{2N}_{\frac{(k+1)T}{2N}}) - F(\widetilde{Y}^{2N}_{\frac{kT}{2N}})\right) + \delta_{2N} h(\widetilde{Y}^{2N}_{\frac{kT}{2N}}) + \sqrt{1-\rho^2}$$
$$\times\sqrt{\left(\psi(\widetilde{Y}^{2N}_{\frac{kT}{2N}}) + \frac{\sigma\psi'(\widetilde{Y}^{2N}_{\frac{kT}{2N}})}{\delta_{2N}}\int_{\frac{kT}{2N}}^{\frac{(k+1)T}{2N}}(W_s - W_{\frac{kT}{2N}})ds\right) \vee \underline{\psi}}\ \left(B_{\frac{(k+1)T}{2N}} - B_{\frac{kT}{2N}}\right)$$

Denote by $v_k^{2N} = \sqrt{1-\rho^2}\sqrt{\left(\psi(\widetilde{Y}^{2N}_{\frac{kT}{2N}}) + \frac{\sigma\psi'(\widetilde{Y}^{2N}_{\frac{kT}{2N}})}{\delta_{2N}}\int_{\frac{kT}{2N}}^{\frac{(k+1)T}{2N}}(W_s - W_{\frac{kT}{2N}})ds\right) \vee \underline{\psi}}$ *the random variable which multiplies the increment of the Brownian motion* $\left(B_{\frac{(k+1)T}{2N}} - B_{\frac{kT}{2N}}\right)$*. Because of the independence properties,* $\left(\widetilde{X}^N_{t_k}\right)_{0\leq k\leq N}$ *has the same distribution law as the vector* $\left(\widetilde{\widetilde{X}}^N_{t_k}\right)_{0\leq k\leq N}$ *defined inductively by* $\widetilde{\widetilde{X}}^N_{t_0} = \log(s_0)$ *and*

$$\forall 0 \leq k \leq N-1,\ \widetilde{\widetilde{X}}^N_{t_{k+1}} = \widetilde{\widetilde{X}}^N_{t_k} + \rho\left(F(\widetilde{Y}^N_{t_{k+1}}) - F(\widetilde{Y}^N_{t_k})\right) + \delta_N h(\widetilde{Y}^N_{t_k})$$
$$+\sqrt{1-\rho^2}\sqrt{\left(\psi(\widetilde{Y}^N_{t_k}) + \frac{\sigma\psi'(\widetilde{Y}^N_{t_k})}{\delta_N}\int_{t_k}^{t_{k+1}}(W_s - W_{t_k})ds\right) \vee \underline{\psi}}\ \Delta\widetilde{B}^N_{k+1}$$

where

$$\Delta\widetilde{B}^N_{k+1} = \sqrt{2}\left[\frac{v_{2k}^{2N}\left(B_{\frac{(2k+1)T}{2N}} - B_{\frac{2kT}{2N}}\right) + v_{2k+1}^{2N}\left(B_{\frac{(2k+2)T}{2N}} - B_{\frac{(2k+1)T}{2N}}\right)}{\sqrt{\left(v_{2k}^{2N}\right)^2 + \left(v_{2k+1}^{2N}\right)^2}}\right]$$

Going over the proof of the theorem, one can show in the same way that

$$\mathbb{E}\left[\max_{0\leq k\leq N}\left|\widetilde{\widetilde{X}}^N_{t_k} - \widetilde{X}^{2N}_{t_k}\right|^2\right] = \mathcal{O}(N^{-2}) \tag{2.10}$$

Hence, one can apply the multilevel Monte Carlo method to compute the expectation of a Lipschitz continuous functional of X *and reduce the computational cost to achieve a desired root-mean-square error of* $\epsilon > 0$ *to a* $\mathcal{O}(\epsilon^{-2})$*.*

As a matter of fact, the particular structure of our scheme enabled us to reconstruct the coupling which allows to efficiently control the error between the scheme with time step $\frac{T}{N}$ *and the one with time step* $\frac{T}{2N}$*. This does not seem possible with the Cruzeiro et al. [27] scheme.*

From a practical point of view, it is more interesting to obtain a convergence result for the stock price. It is also more challenging because the exponential function is not globally Lipschitz continuous. We can nevertheless state the following corollary with some general assumptions and we will see in the next section that we can make them more precise in case $(Y_t)_{t\in[0,T]}$ is an Ornstein-Uhlenbeck process.

Corollary 24 — *Let $p \geq 1$. Under the assumptions of Theorem 20 and if*
($\mathcal{H}7$)

$$\exists \epsilon > 0 \text{ such that } \mathbb{E}\left[\max_{0\leq k\leq N} S_{t_k}^{2p+\epsilon}\right] + \mathbb{E}\left[\max_{0\leq k\leq N} e^{(2p+\epsilon)\widetilde{X}_{t_k}^N}\right] < \infty$$

then there exists a positive constant C independent of N such that

$$\mathbb{E}\left[\max_{0\leq k\leq N}\left|e^{\widetilde{X}_{t_k}} - e^{\widetilde{X}_{t_k}^N}\right|^{2p}\right] \leq \frac{C}{N^{2p}}$$

Proof : Using Hölder inequality we have that

$$\begin{aligned}
\mathbb{E}\left[\max_{0\leq k\leq N}\left|e^{\widetilde{X}_{t_k}} - e^{\widetilde{X}_{t_k}^N}\right|^{2p}\right] &\leq \mathbb{E}\left[\max_{0\leq k\leq N}\left(e^{2p\widetilde{X}_{t_k}} \vee e^{2p\widetilde{X}_{t_k}^N}\right)\left|\widetilde{X}_{t_k} - \widetilde{X}_{t_k}^N\right|^{2p}\right] \\
&\leq \left(\mathbb{E}\left[\max_{0\leq k\leq N} S_{t_k}^{2p+\epsilon}\right] + \mathbb{E}\left[\max_{0\leq k\leq N} e^{(2p+\epsilon)\widetilde{X}_{t_k}^N}\right]\right)^{\frac{2p}{2p+\epsilon}} \\
&\quad \times \left(\mathbb{E}\left[\max_{0\leq k\leq N}\left|\widetilde{X}_{t_k} - \widetilde{X}_{t_k}^N\right|^{\frac{2p\epsilon+4p^2}{\epsilon}}\right]\right)^{\frac{\epsilon}{2p+\epsilon}}
\end{aligned}$$

We conclude by assumption ($\mathcal{H}7$) and Theorem 20. □

Remark 25 — *Had we introduced a new cut-off to our scheme as follows*

$$\begin{aligned}
\widetilde{X}_{t_{k+1}}^N &= \widetilde{X}_{t_k}^N + \rho\left(F(\widetilde{Y}_{t_{k+1}}^N) - F(\widetilde{Y}_{t_k}^N)\right) + \delta_N h(\widetilde{Y}_{t_k}^N) \\
&\quad + \sqrt{1-\rho^2}\sqrt{\left(\psi(\widetilde{Y}_{t_k}^N) + \frac{\sigma\psi'(\widetilde{Y}_{t_k}^N)}{\delta_N}\int_{t_k}^{t_{k+1}}(W_s - W_{t_k})ds\right) \wedge \overline{\psi} \vee \underline{\psi}}\ \Delta B_{k+1}
\end{aligned}$$

assumption ($\mathcal{H}7$) would have been induced by assuming that the functions F, f and h are bounded.

2.1.2 Special case of an Ornstein-Uhlenbeck process driving the volatility

For many stochastic volatility models, the process $(Y_t)_{t\in[0,T]}$ which drives the volatility is an Ornstein-Uhlenbeck process. For example, this is the case for all the models cited in the introduction but the Heston model. Therefore, it is useful to focus on this particular case. We will hereafter suppose that $(Y_t)_{t\in[0,T]}$ is solution of the following SDE

$$dY_t = \nu dW_t + \kappa(\theta - Y_t)dt \tag{2.11}$$

with ν, κ and θ three positive constants. Since exact simulation is possible, we can replace the Milstein discretization by the true solution in our previous scheme :

WeakTraj_1 scheme when Y is an O-U process

$$\widetilde{X}^N_{t_{k+1}} = \widetilde{X}^N_{t_k} + \rho\left(F(Y_{t_{k+1}}) - F(Y_{t_k})\right) + \delta_N h(Y_{t_k}) + \sqrt{1-\rho^2}\sqrt{\left(\psi(Y_{t_k}) + \frac{\nu\psi'(Y_{t_k})}{\delta_N}\int_{t_k}^{t_{k+1}}(W_s - W_{t_k})ds\right) \vee \underline{\psi}}\,\Delta B_{k+1} \tag{2.12}$$

Note that we require the exact simulation of both $(Y_{t_k}, Y_{t_{k+1}})$ and $\int_{t_k}^{t_{k+1}}(W_s - W_{t_k})ds$. The unique solution of (2.11) is $Y_t = y_0 e^{-\kappa t} + \theta(1-e^{-\kappa t}) + \nu\int_0^t e^{-\kappa(t-s)}dW_s$ and one can easily deduce that, $\forall k \in \{0,\dots,N-1\}$, $\begin{pmatrix} Y_{t_{k+1}} - e^{-\kappa\delta_N}Y_{t_k} \\ \int_{t_k}^{t_{k+1}}(W_s - W_{t_k})ds \end{pmatrix} \sim \mathcal{N}(M,\Gamma)$ where $M = \begin{pmatrix} \theta(1-e^{-\kappa\delta_N}) \\ 0 \end{pmatrix}$ and $\Gamma = \begin{pmatrix} \frac{\nu^2}{2\kappa}(1-e^{-2\kappa\delta_N}) & \frac{\nu}{\kappa^2}(1-e^{-\kappa\delta_N}(1+\kappa\delta_N)) \\ \frac{\nu}{\kappa^2}(1-e^{-\kappa\delta_N}(1+\kappa\delta_N)) & \frac{\delta_N^3}{3} \end{pmatrix}$.

We first state the following technical lemma whose proof is postponed to the appendix :

Lemma 26 — $\forall\, c_1 > 0, c_2 \in [0,1)$,

$$\mathbb{E}\left(e^{c_1 \sup_{0\le t\le T}|Y_t|^{1+c_2}}\right) < \infty.$$

As might be expected, it is possible to weaken the assumptions of Theorem 20. In particular, we relax the assumption on the lower bound of the volatility ($\mathcal{H}2$) and replace it with a weaker one (see assumption ($\mathcal{H}10$) below).

Theorem 27 — *Let $p \ge 1$. Suppose that Y is solution of (2.11) and that the scheme is defined by (2.12). Under assumption ($\mathcal{H}2$) of Theorem 20 and if*

($\mathcal{H}8$) *f is a $\mathcal{C}^3$ function*

($\mathcal{H}9$) *there exist three constants $c_0 > 0, c_1 > 0$ and $c_2 \in [0,1)$ such that, $\forall y \in \mathbb{R}$,*

$$\begin{aligned}
&\left|\kappa(\theta - y)h'(y) + \frac{\nu^2}{2}h''(y)\right| \le c_0 e^{c_1|y|^{1+c_2}} \\
&\left|h'(y)\right| \le c_0 e^{c_1|y|^{1+c_2}} \\
&\left|\kappa(\theta - y)\psi'(y) + \frac{\nu^2}{2}\psi''(y)\right| \le c_0 e^{c_1|y|^{1+c_2}} \\
&\left|\psi''(y)\right| \le c_0 e^{c_1|y|^{1+c_2}}
\end{aligned}$$

then there exists a constant C independent of the number of time steps N such that

$$\mathbb{E}\left[\max_{0\le k\le N}\left|\widetilde{X}_{t_k} - \widetilde{X}^N_{t_k}\right|^{2p}\right] \le \frac{C}{N^{2p}}$$

The same result holds true when we replace assumption ($\mathcal{H}2$) by

($\mathcal{H}$10) *There exist two positive constants C and ϵ such that $\forall y \in \mathbb{R}$,*

$$\begin{aligned}
\psi(y) &> 0 \\
|\psi'(y)| &\leq C\psi(y) \\
\sup_{t\leq T} \mathbb{E}\left(\psi^{p(1+\epsilon)}(Y_t)\right) &< \infty \\
\sup_{t\leq T} \mathbb{E}\left(\frac{1}{\psi^{p(1+\epsilon)}(Y_t)}\right) &< \infty.
\end{aligned}$$

Proof : The proof of the first part of the theorem repeats the proof of Theorem 20 with fewer terms to control because of the exact simulation of $(Y_t)_{t\in[0,T]}$. At the places where we used assumptions ($\mathcal{H}$3) and ($\mathcal{H}$4), we use assumption ($\mathcal{H}$9) together with Lemma 26.

We now focus on the second part of the theorem. According to equation (2.8), all we have to show is the existence of a positive constant C independent of N such that $\forall j \in \{0,\dots,N-1\}$

$$\mathbb{E}\left[\left|\sqrt{\frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds} - \sqrt{\left(\psi(Y_{t_j}) + \frac{\nu\psi'(Y_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s - W_{t_j})ds\right)\vee\underline{\psi}}\right|^{2p}\right] \leq \frac{C}{N^{2p}}$$

We will adopt the following notations

- $A_j = \frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds$
- $D_j = \left(\psi(Y_{t_j}) + \frac{\nu\psi'(Y_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s - W_{t_j})ds\right)\vee\underline{\psi}$

Thanks to assumption ($\mathcal{H}$10), we have that $\forall j \in \{0,\dots,N-1\}, A_j > 0$ and $D_j \geq 0$. The idea of the proof is to isolate the case where D_j is small which is problematic since the square root is not Lipschitz continuous in the neighborhood of 0 :

$$\begin{aligned}
\left|\sqrt{A_j} - \sqrt{D_j}\right|^{2p} &= \left|\sqrt{A_j} - \sqrt{D_j}\right|^{2p}\mathbb{1}_{\{D_j\leq\psi(Y_{t_j})/2\}} + \left|\sqrt{A_j} - \sqrt{D_j}\right|^{2p}\mathbb{1}_{\{D_j>\psi(Y_{t_j})/2\}} \\
&\leq 2^{2p-1}\left(A_j^p + \frac{\psi^p(Y_{t_j})}{2^p}\right)\mathbb{1}_{\{D_j\leq\psi(Y_{t_j})/2\}} \\
&\quad + 2^{2p-2}\left(\frac{1}{A_j^p} + \frac{2^p}{\psi^p(Y_{t_j})}\right)|A_j - D_j|^{2p}\mathbb{1}_{\{D_j>\psi(Y_{t_j})/2\}}
\end{aligned}$$

We take the expectation and apply Hölder inequality to obtain

$$\mathbb{E}\left[\left|\sqrt{A_j} - \sqrt{D_j}\right|^{2p}\right] \leq C(\epsilon_1 + \epsilon_2)$$

with

$$\epsilon_1 = \left(\mathbb{E}\left[\left(A_j^p + \frac{\psi^p(Y_{t_j})}{2^p}\right)^{1+\epsilon}\right]\right)^{\frac{1}{1+\epsilon}}\left(\mathbb{P}\left(D_j \leq \frac{\psi(Y_{t_j})}{2}\right)\right)^{\frac{\epsilon}{1+\epsilon}}$$

and

$$\epsilon_2 = \left(\mathbb{E}\left[\left(\frac{1}{A_j^p} + \frac{2^p}{\psi^p(Y_{t_j})}\right)^{1+\epsilon}\right]\right)^{\frac{1}{1+\epsilon}}\left(\mathbb{E}\left[|A_j - D_j|^{2p\frac{1+\epsilon}{\epsilon}}\right]\right)^{\frac{\epsilon}{1+\epsilon}}.$$

Let us begin with the second term. Following the estimation of $\overline{I}_2^j$ in the proof of Theorem 20, we show that

$$\begin{aligned}\mathbb{E}\left[|A_j - D_j|^{2p\frac{1+\epsilon}{\epsilon}}\right] &= \mathbb{E}\left[\left|\frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds - \left(\psi(Y_{t_j}) + \frac{\nu\psi'(Y_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s - W_{t_j})ds\right)\vee \underline{\psi}\right|^{2p\frac{1+\epsilon}{\epsilon}}\right] \\ &\leq C\delta_N^{2p\frac{1+\epsilon}{\epsilon}}\end{aligned}$$

Thanks to assumption ($\mathcal{H}10$) and Jensen's inequality, we also have that

$$\begin{aligned}\left(\mathbb{E}\left[\left(\frac{1}{A_j^p} + \frac{2^p}{\psi^p(Y_{t_j})}\right)^{1+\epsilon}\right]\right)^{\frac{1}{1+\epsilon}} &\leq 2^{\frac{\epsilon}{1+\epsilon}}\left(\frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\mathbb{E}\left(\frac{1}{\psi^{p(1+\epsilon)}(Y_s)}\right)ds + 2^{p(1+\epsilon)}\mathbb{E}\left(\frac{1}{\psi^{p(1+\epsilon)}(Y_{t_j})}\right)\right)^{\frac{1}{1+\epsilon}} \\ &\leq C\end{aligned}$$

Hence $\epsilon_2 \leq \frac{C}{N^{2p}}$. Now let us turn to ϵ_1. Note first that assumption ($\mathcal{H}10$) enables us to show that there exists a positive constant C independent of N such that $\left(\mathbb{E}\left[\left(A_j^p + \frac{\psi^p(Y_{t_j})}{2^p}\right)^{1+\epsilon}\right]\right)^{\frac{1}{1+\epsilon}} \leq C$. Finally, what is left to prove is that $\mathbb{P}\left(D_j \leq \frac{\psi(Y_{t_j})}{2}\right) \leq \frac{C}{N^{2p\frac{1+\epsilon}{\epsilon}}}$. In fact, we can show that $\forall\alpha > 0, \exists C_\alpha > 0$ such that $\mathbb{P}\left(D_j \leq \frac{\psi(Y_{t_j})}{2}\right) \leq \frac{C_\alpha}{N^\alpha}$:

$$\begin{aligned}\mathbb{P}\left(D_j \leq \frac{\psi(Y_{t_j})}{2}\right) &\leq \mathbb{P}\left(\frac{\nu\psi'(Y_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s - W_{t_j})ds \leq -\frac{\psi(Y_{t_j})}{2}\right) \\ &= \mathbb{P}\left(|G| \geq \frac{\sqrt{3}\psi(Y_{t_j})}{2\sqrt{\delta_N}\nu|\psi'(Y_{t_j})|}\right)\end{aligned}$$

where G is a centered reduced Gaussian random variable independent of Y_{t_j}.

Thanks to assumption ($\mathcal{H}10$), $\exists C > 0$ s.t. $\mathbb{P}\left(|G| \geq \frac{\sqrt{3}\psi(Y_{t_j})}{2\sqrt{\delta_N}\nu|\psi'(Y_{t_j})|}\right) \leq 2\mathbb{P}\left(G \geq \frac{C}{\sqrt{\delta_N}}\right)$ and using the following standard upper bound of the Gaussian tail probability : $\forall t > 0, \mathbb{P}(G \geq t) \leq \frac{e^{-\frac{t^2}{2}}}{t\sqrt{2\pi}}$, we conclude. □

Remark 28 —

- *The fact that we can simulate exactly the volatility process without affecting the order of convergence of the scheme is yet another advantage of our approach over the Cruzeiro et al. [27] scheme. On the other hand, the Kahl and Jäckel [61] scheme allows the exact simulation of $(Y_t)_{t\in[0,T]}$. Applied to the SDE (2.3), it writes as*

$$\begin{aligned}X_{t_{k+1}}^{IJK} = X_{t_k}^{IJK} &+ \left(r - \frac{f^2(Y_{t_{k+1}}) + f^2(Y_{t_k})}{4}\right)\delta_N + \rho f(Y_{t_k})\Delta W_{k+1} \\ &+ \sqrt{1-\rho^2}\frac{f(Y_{t_{k+1}}) + f(Y_{t_k})}{2}\Delta B_{k+1} + \frac{\rho\nu}{2}f'(Y_{t_k})\left((\Delta W_{k+1})^2 - \delta_N\right)\end{aligned} \tag{2.13}$$

Note that it is close to our scheme insofar as it takes advantage of the structure of the SDE (for example, unlike the Cruzeiro et al. [27] scheme, it allows the use of the coupling introduced in Remark 23). The main difference, which explains why our scheme has better weak trajectorial convergence order, is that we discretize more accurately the integral of $f(Y_t)$ with respect to the Brownian motion $(B_t)_{t\in[0,T]}$. If, instead of a trapezoidal method, one uses the same discretization as for the WeakTraj_1 scheme, then it can be shown that this modified IJK scheme will exhibit a first order weak trajectorial convergence.

- *One can easily check that this theorem applies for the Scott [112] model (and therefore for the Hull and White [57] model) where we have $h(y) = r - \frac{e^{2y}}{2} - \rho e^y(\frac{\kappa}{\nu}(\theta - y) + \frac{\nu}{2})$ and $\psi(y) = e^{2y}$. The Stein and Stein [115] and the quadratic Gaussian models do not satisfy the assumption $|\psi'(y)| \leq C\psi(y)$.*
- *It is possible to improve the convergence at fixed times up to the order $\frac{3}{2}$. Following Lapeyre and Temam [81] who approximate an integral of the form $\int_{t_k}^{t_{k+1}} g(Y_s)ds$ for a twice differentiable function g by $\delta_N g(Y_{t_k}) + \nu g'(Y_{t_k}) \int_{t_k}^{t_{k+1}} (W_s - W_{t_k})ds + (\kappa(\theta - Y_{t_k})g'(Y_{t_k}) + \frac{\nu^2}{2} g''(Y_{t_k}))\frac{\delta_N^2}{2}$, we obtain the following scheme*

OU_Improved scheme

$$\widetilde{X}^N_{t_{k+1}} = \widetilde{X}^N_{t_k} + \rho\left(F(Y_{t_{k+1}}) - F(Y_{t_k})\right) + \widetilde{h}_k + \sqrt{1-\rho^2}\sqrt{\widetilde{\psi}_k}\ \Delta B_{k+1} \tag{2.14}$$

where $\widetilde{h}_k = \delta_N h(Y_{t_k}) + \nu h'(Y_{t_k}) \int_{t_k}^{t_{k+1}} (W_s - W_{t_k})ds + (\kappa(\theta - Y_{t_k})h'(Y_{t_k}) + \frac{\nu^2}{2} h''(Y_{t_k}))\frac{\delta_N^2}{2}$ and $\widetilde{\psi}_k = \left(\psi(Y_{t_k}) + \frac{\nu\psi'(Y_{t_k})}{\delta_N} \int_{t_k}^{t_{k+1}} (W_s - W_{t_k})ds + (\kappa(\theta - Y_{t_k})\psi'(Y_{t_k}) + \frac{\nu^2}{2}\psi''(Y_{t_k}))\frac{\delta_N}{2}\right) \vee \underline{\psi}$. Mimicking the proof of Theorem 20, one can show that

$$\max_{0\leq k\leq N} \mathbb{E}\left[\left|\widehat{X}_{t_k} - \widehat{X}^N_{t_{k+1}}\right|^2\right] = \mathcal{O}\left(N^{-3}\right)$$

where $\widehat{X}_{t_k}$ and $\widehat{X}^N_{t_{k+1}}$ have respectively the same distribution as X_{t_k} and $\widetilde{X}^N_{t_k}$:

$$\widehat{X}_{t_k} = X_0 + \rho(F(Y_{t_k}) - F(y_0)) + \int_0^{t_k} h(Y_s)ds + \sqrt{1-\rho^2}\sqrt{\frac{1}{t_k}\int_0^{t_k} \psi(Y_s)ds}\ B_{t_k}$$

and

$$\widehat{X}^N_{t_k} = X_0 + \rho\left(F(Y_{t_k}) - F(y_0)\right) + \sum_{j=0}^{k-1} \widetilde{h}_j + \sqrt{1-\rho^2}\sqrt{\frac{\delta_N}{t_k}\sum_{j=0}^{k-1} \widetilde{\psi}_j}\ B_{t_k}.$$

As for the stock, we can prove the same convergence result under some additional assumptions which are more explicit than assumption ($\mathcal{H}7$) of Corollary 24. To do so, let us make the following changes in our scheme so that we can control its exponential moments :

$$\begin{aligned}\widetilde{X}^N_{t_{k+1}} &= \widetilde{X}^N_{t_k} + \rho\left(F(Y_{t_{k+1}}) - F(Y_{t_k})\right) + \delta_N h(Y_{t_k}) \\ &+ \sqrt{1-\rho^2}\sqrt{\left(\psi(Y_{t_k}) + \frac{\nu\psi'(Y_{t_k})}{\delta_N}\int_{t_k}^{t_{k+1}} (W_s - W_{t_k})ds\right) \wedge \widehat{\psi}(Y_{t_k}) \vee \underline{\psi}}\ \Delta B_{k+1}\end{aligned} \tag{2.15}$$

Proposition 29 — *Suppose that Y is solution of (2.11) and that the scheme is defined by (2.15). Under the assumptions ($\mathcal{H}8$), ($\mathcal{H}9$) and ($\mathcal{H}10$) of Theorem 27 and if*

($\mathcal{H}11$) *there exists $\beta \in (0,1)$ and $K > 0$ such that $\forall y \in \mathbb{R}$*

$$\begin{aligned} |h(y)| + |F(y)| + |f'(y)| &\leq K(1+|y|^{1+\beta}) \\ |f(y)| &\leq K(1+|y|^{\beta}) \end{aligned}$$

then, $\forall p \geq 1$, there exists a positive constant C independent of N such that

$$\mathbb{E}\left[\max_{0\leq k\leq N}\left|e^{\widetilde{X}_{t_k}} - e^{\widetilde{X}^N_{t_k}}\right|^{2p}\right] \leq \frac{C}{N^{2p}}.$$

The same result holds true if one replaces assumption ($\mathcal{H}10$) by assumption ($\mathcal{H}2$) together with the assumption that $\exists C > 0$ for which $\forall y \in \mathbb{R}, |\psi'(y)| \leq C\psi(y)$.

Proof : We go over the proof of Corollary 24. The fact that $\mathbb{E}\left[\max_{0\leq k\leq N}\left|\widetilde{X}_{t_k} - \widetilde{X}^N_{t_k}\right|^{4p}\right] = \mathcal{O}(\frac{1}{N^{4p}})$ is not a straightforward consequence of Theorem 27 anymore because we have introduced some changes in our scheme. However, looking through the proof of the theorem, one can see that it is enough to prove the following inequality : $\forall j \in \{0,\dots,N-1\}$

$$\mathbb{E}\left[\left|\sqrt{\frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds} - \sqrt{\left(\psi(Y_{t_j}) + \frac{\nu\psi'(Y_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s - W_{t_j})ds\right)\wedge\widehat{\psi}(Y_{t_j})\vee\underline{\psi}}\right|^{2p}\right] \leq \frac{C}{N^{2p}} \tag{2.16}$$

When $\overline{\psi}$ is finite, since $\frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds$ is smaller than $\widehat{\psi}(Y_{t_k}) = \overline{\psi}$, we can remove the new cut-off from the left hand side of (2.16) and then proceed like in Theorem 27. When $\overline{\psi} = +\infty$, on the event $\left(\psi(Y_{t_j}) + \frac{\nu\psi'(Y_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s - W_{t_j})ds\right) \leq \widehat{\psi}(Y_{t_j})$, we recover our original scheme and we prove (2.16) like in Theorem 27. Then, using the Gaussian arguments developed in the end of the proof of Theorem 27, we control the probability of the complementary event to conclude.

Now, what is left to prove is that assumption ($\mathcal{H}7$) is satisfied. On the one hand, we have that

$$\begin{aligned} \mathbb{E}\left[\max_{0\leq k\leq N} S^{4p}_{t_k}\right] &= \mathbb{E}\left[\max_{0\leq k\leq N}\left(S_0 + \int_0^{t_k} rS_s ds + \int_0^{t_k} f(Y_s)S_s\left(\rho dW_s + \sqrt{1-\rho^2}dB_s\right)\right)^{4p}\right] \\ &\leq C\left(1 + \int_0^T \mathbb{E}\left(S_t^{4p}(1+f^{4p}(Y_t))\right)dt\right) \\ &\leq C\left(1 + \int_0^T \sqrt{\mathbb{E}(S_t^{8p})}\sqrt{\mathbb{E}\left((1+f^{4p}(Y_t))^2\right)}dt\right) \end{aligned}$$

Thanks to assumption ($\mathcal{H}11$) and Lemma 26, there exists $C > 0$ such that $\sqrt{\mathbb{E}\left((1+f^{4p}(Y_t))^2\right)} \leq C$. Observe that conditionally on $(Y_t)_{t\in[0,T]}$,

$$X_t \sim \mathcal{N}\left(\log(s_0) + \rho(F(Y_t) - F(y_0)) + \int_0^t h(Y_s)ds\,,\,(1-\rho^2)\int_0^t f^2(Y_s)ds\right) \tag{2.17}$$

so, by Jensen's inequality and assumption ($\mathcal{H}11$)

$$\begin{aligned}
\mathbb{E}\left(S_t^{8p}\right) &= \mathbb{E}\left(e^{8p\left(\log(s_0)+\rho(F(Y_t)-F(y_0))+\int_0^t h(Y_s)ds\right)}e^{32p^2(1-\rho^2)\int_0^t f^2(Y_s)ds}\right) \\
&\leq \mathbb{E}\left(e^{8p(\log(s_0)+\rho(F(Y_t)-F(y_0)))}\frac{1}{t}\int_0^t e^{t\left(8ph(Y_s)+32p^2(1-\rho^2)f^2(Y_s)\right)}ds\right) \\
&\leq C\mathbb{E}\left(e^{C\sup_{0\leq t\leq T}|Y_t|^{1+\beta}}\right)
\end{aligned}$$

Using Lemma 26, we deduce that $\mathbb{E}\left[\max_{0\leq k\leq N} S_{t_k}^{4p}\right] < \infty$.

On the other hand, using Cauchy-Schwartz inequality, we have that

$$\begin{aligned}
\mathbb{E}\left[\max_{0\leq k\leq N} e^{4p\widetilde{X}_{t_k}^N}\right] &= \mathbb{E}\left[\max_{0\leq k\leq N}\exp\left(4p\left(X_0+\rho(F(Y_{t_k})-F(y_0))+\sum_{j=0}^{k-1}\delta_N h(Y_{t_j})+\sum_{j=0}^{k-1}\sqrt{1-\rho^2}\right.\right.\right. \\
&\quad \left.\left.\left.\times\sqrt{\left(\psi(Y_{t_j})+\frac{\nu\psi'(Y_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right)\wedge\widehat{\psi}(Y_{t_j})\vee\underline{\psi}}\,\Delta B_{j+1}\right)\right)\right] \\
&\leq \sqrt{\widetilde{E}_1^N}\sqrt{\widetilde{E}_2^N}
\end{aligned}$$

where

$$\widetilde{E}_1^N = \mathbb{E}\left[\max_{0\leq k\leq N} e^{8p\left(X_0+\rho(F(Y_{t_k})-F(y_0))+\sum_{j=0}^{k-1}\delta_N h(Y_{t_j})\right)}\right]$$

and

$$\widetilde{E}_2^N = \mathbb{E}\left[\max_{0\leq k\leq N} e^{8p\sqrt{1-\rho^2}\sum_{j=0}^{k-1}\sqrt{\left(\psi(Y_{t_j})+\frac{\nu\psi'(Y_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right)\wedge\widehat{\psi}(Y_{t_j})\vee\underline{\psi}}\,\Delta B_{j+1}}\right].$$

Using the same argument as before, we show that $\widetilde{E}_1^N \leq C\mathbb{E}\left(e^{C\sup_{0\leq t\leq T}|Y_t|^{1+\beta}}\right) < \infty$.

Denote by $D_j = \left(\psi(Y_{t_j})+\frac{\sigma\psi'(Y_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right)\wedge\widehat{\psi}(Y_{t_j})\vee\underline{\psi}$. Using Doob's maximal inequality for the positive submartingale $\left(e^{4p\sqrt{1-\rho^2}\sum_{j=0}^{k-1}\sqrt{D_j}\Delta B_{j+1}}\right)_{0\leq k\leq N}$ (see Theorem 3.8 p. 13 of Karatzas and Shreve [63] for example), we also have that

$$\begin{aligned}
\widetilde{E}_2^N &\leq 4\mathbb{E}\left(e^{8p\sqrt{1-\rho^2}\sum_{j=0}^{N-1}\sqrt{D_j}\,\Delta B_{j+1}}\right) \\
&= 4\mathbb{E}\left(\prod_{j=0}^{N-1} e^{32p^2\delta_N(1-\rho^2)D_j}\right) \\
&\leq 4\mathbb{E}\left(\max_{0\leq k\leq N-1} e^{32p^2(1-\rho^2)\widehat{\psi}(Y_{t_j})}\right)
\end{aligned}$$

By virtue of assumption ($\mathcal{H}11$), $\widetilde{E}_2^N < \infty$ which concludes the proof. □

2.2 A second order weak scheme

Integrating the first stochastic differential equation in (2.4) gives

$$X_t = \log(s_0) + \rho(F(Y_t) - F(y_0)) + \int_0^t h(Y_s)ds + \sqrt{1-\rho^2}\int_0^t f(Y_s)dB_s \tag{2.18}$$

We are only left with an integral with respect to time which can be handled by the use of a trapezoidal scheme and a stochastic integral where the integrand is independent of the Brownian motion. Hence, conditionally on $(Y_t)_{t\in[0,T]}$,

$$X_T \sim \mathcal{N}\left(\log(s_0) + \rho(F(Y_T) - F(y_0)) + m_T\,,\,(1-\rho^2)v_T\right) \tag{2.19}$$

where $m_T = \int_0^T h(Y_s)ds$ and $v_T = \int_0^T f^2(Y_s)ds$. This suggests that, in order to properly approximate the law of X_T, one should accurately approximate the law of Y_T and carefully handle integrals with respect to time of functions of the process $(Y_t)_{t\in[0,T]}$. We thus define our weak scheme as follows

Weak_2 scheme

$$\overline{X}_T^N = \log(s_0) + \rho(F(\overline{Y}_T^N) - F(y_0)) + \overline{m}_T^N + \sqrt{(1-\rho^2)\overline{v}_T^N}G \tag{2.20}$$

where $\overline{m}_T^N = \delta_N \sum_{k=0}^{N-1} \frac{h(\overline{Y}_{t_k}^N)+h(\overline{Y}_{t_{k+1}}^N)}{2}$, $\overline{v}_T^N = \delta_N \sum_{k=0}^{N-1} \frac{f^2(\overline{Y}_{t_k}^N)+f^2(\overline{Y}_{t_{k+1}}^N)}{2}$, $(\overline{Y}_{t_k}^N)_{0\le k\le N}$ is the Ninomiya-Victoir scheme of $(Y_t)_{t\in[0,T]}$ and G is an independent centered reduced Gaussian random variable. Note that, conditionally on $(\overline{Y}_{t_k}^N)_{0\le k\le N}$, $\overline{X}_t^N$ is also a Gaussian random variable with mean $\log(s_0) + \rho(F(\overline{Y}_T^N) - F(y_0)) + \overline{m}_T^N$ and variance $(1-\rho^2)\overline{v}_T^N$.

It is well known that the Ninomiya and Victoir [98] scheme is of weak order two. For the sake of completeness, we give its definition in our setting :

$$\begin{cases} \overline{Y}_0^N = y_0 \\ \forall 0 \le k \le N-1, \overline{Y}_{t_{k+1}}^N = \exp\left(\frac{T}{2N}V_0\right)\exp\left((W_{t_{k+1}} - W_{t_k})V\right)\exp\left(\frac{T}{2N}V_0\right)(\overline{Y}_{t_k}^N) \end{cases}$$

where $V_0 : x \mapsto b(x) - \frac{1}{2}\sigma\sigma'(x)$ and $V : x \mapsto \sigma(x)$. The notation $\exp(tV)(x)$ stands for the solution, at time t and starting from x, of the ODE $\eta'(t) = V(\eta(t))$. What is nice with our setting is that we are in dimension one and thus such ODEs can be solved explicitly. Indeed, if ζ is a primitive of $\frac{1}{V}$: $\zeta(t) = \int_0^t \frac{1}{V(s)}ds$, then the solution writes as $\eta(t) = \zeta^{-1}(t + \zeta(x))$.

Note that our scheme can be seen as a splitting scheme for the SDE satisfied by $(Z_t = X_t - \rho F(Y_t), Y_t)$:

$$\begin{cases} dZ_t &= h(Y_t)dt + \sqrt{1-\rho^2}f(Y_t)dB_t \\ dY_t &= b(Y_t)dt + \sigma(Y_t)dW_t \end{cases} \tag{2.21}$$

The differential operator associated to (2.21) writes as

$$\mathcal{L}v(z,y) = h(y)\frac{\partial v}{\partial z} + b(y)\frac{\partial v}{\partial y} + \frac{\sigma^2(y)}{2}\frac{\partial^2 v}{\partial y^2} + \frac{(1-\rho^2)}{2}f^2(y)\frac{\partial^2 v}{\partial z^2} = \mathcal{L}_Y v(z,y) + \mathcal{L}_Z v(z,y)$$

where $\mathcal{L}_Y v(z,y) = b(y)\frac{\partial v}{\partial y} + \frac{\sigma^2(y)}{2}\frac{\partial^2 v}{\partial y^2}$ and $\mathcal{L}_Z v(z,y) = h(y)\frac{\partial v}{\partial z} + \frac{(1-\rho^2)}{2} f^2(y)\frac{\partial^2 v}{\partial z^2}$. One can check that our scheme amounts to first integrate exactly $\mathcal{L}_Z$ over a half time step then apply the Ninomiya-Victoir scheme to $\mathcal{L}_Y$ over a time step and finally integrate exactly $\mathcal{L}_Z$ over a half time step. According to results on splitting (see Alfonsi [2] or Tanaka and Kohatsu-Higa [119] for example) one expects this scheme to exhibit second order weak convergence. We will not use this point of view to prove our convergence result stated in the next theorem, since we need to apply test functions with exponential growth to X_T to be able to analyse weak convergence of the stock price.

Theorem 30 — *Suppose that $\rho \in (-1,1)$. If the following assumptions hold*

(H12) *b and σ are respectively $\mathcal{C}^4$ and $\mathcal{C}^5$, with bounded derivatives of any order greater or equal to 1.*

(H13) *h and f are $\mathcal{C}^4$ and F is $\mathcal{C}^6$. The three functions are bounded together with all their derivatives.*

(H14) $\underline{\psi} > 0$

then, for any measurable function g verifying $\exists c \geq 0, \mu \in [0,2)$ such that $\forall x \in \mathbb{R}, |g(x)| \leq ce^{|x|^\mu}$, there exists $C > 0$ such that

$$\left|\mathbb{E}\Big(g(X_T)\Big) - \mathbb{E}\left(g(\overline{X}_T^N)\right)\right| \leq \frac{C}{N^2}$$

In terms of the asset price, we easily deduce the following corollary :

Corollary 31 — *Under the assumptions of Theorem 30, for any measurable function α verifying $\exists c \geq 0, \mu \in [0,2)$ such that $\forall y > 0, |\alpha(y)| \leq ce^{|\log(y)|^\mu}$, there exists $C > 0$ such that*

$$\left|\mathbb{E}\left(\alpha(S_T)\right) - \mathbb{E}\left(\alpha(e^{\overline{X}_T^N})\right)\right| \leq \frac{C}{N^2}$$

Proof of the theorem : The idea of the proof consists in conditioning by the Brownian motion which drives the volatility process and then applying the weak error analysis of Talay and Tubaro [117].

As stated above, conditionally on $(W_t)_{t\in[0,T]}$, both X_T and $\overline{X}_T^N$ are Gaussian random variables and one can easily show that

$$\begin{aligned}\epsilon &:= \left|\mathbb{E}\left[g(X_T) - g(\overline{X}_T^N)\right]\right| \\ &= \left|\int_{\mathbb{R}} g(x)\mathbb{E}\left[\frac{\exp\left(-\frac{(x-\log(s_0)+\rho F(y_0)-\rho F(Y_T)-m_T)^2}{2(1-\rho^2)v_T}\right)}{\sqrt{2\pi(1-\rho^2)v_T}} - \frac{\exp\left(-\frac{(x-\log(s_0)+\rho F(y_0)-\rho F(\overline{Y}_T^N)-\overline{m}_T^N)^2}{2(1-\rho^2)\overline{v}_T^N}\right)}{\sqrt{2\pi(1-\rho^2)\overline{v}_T^N}}\right] dx\right|\end{aligned}$$

For $x \in \mathbb{R}$, denote by γ_x the function

$$\begin{array}{rcl} \gamma_x : \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+^* & \to & \mathbb{R} \\ (y, m, v) & \mapsto & \dfrac{\exp\left(-\frac{(x-\log(s_0)+\rho F(y_0)-\rho F(y)-m)^2}{2(1-\rho^2)v}\right)}{\sqrt{2\pi(1-\rho^2)v}} \end{array}$$

so that $\epsilon \leq \int_{\mathbb{R}} g(x) \left| \mathbb{E}\left[\gamma_x(Y_T, m_T, v_T) - \gamma_x(\overline{Y}_T^N, \overline{m}_T^N, \overline{v}_T^N)\right]\right| dx$. Consequently, it is enough to show the following intermediate result : $\exists C, K > 0$ and $p \in \mathbb{N}$ such that

$$\forall x \in \mathbb{R}, \left| \mathbb{E}\left[\gamma_x(Y_T, m_T, v_T) - \gamma_x(\overline{Y}_T^N, \overline{m}_T^N, \overline{v}_T^N)\right]\right| \leq \frac{C}{N^2} e^{-Kx^2}(1 + |x|^p). \tag{2.22}$$

We naturally consider the following 3-dimensional degenerate SDE:

$$\left\{ \begin{array}{rcl} dY_t & = & \sigma(Y_t)dW_t + b(Y_t)dt; \quad Y_0 = y_0 \\ dm_t & = & h(Y_t)dt; \quad m_0 = 0 \\ dv_t & = & f^2(Y_t)dt; \quad v_0 = 0 \end{array} \right. \tag{2.23}$$

Note that $(\overline{Y}_T^N, \overline{m}_T^N, \overline{v}_T^N)$ is close to the terminal value of the Ninomiya-Victoir scheme applied to this 3-dimensional SDE. In order to prove (2.22), we need to analyse the dependence of the error on x and not only on N. That is why we resume the error analysis of Ninomiya and Victoir [98] in a more detailed fashion.

For $x \in \mathbb{R}$, let us define the function $u_x : [0, T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+^* \to \mathbb{R}$ by

$$u_x(t, y, m, v) = \mathbb{E}\left[\gamma_x\left((Y_{T-t}, m_{T-t}, v_{T-t})^{(y,m,v)}\right)\right]$$

where we denote by $(Y_{T-t}, m_{T-t}, v_{T-t})^{(y,m,v)}$ the solution at time $T - t$ of (2.23) starting from (y, m, v).

The remainder of the proof leans on the following lemmas. We will use the standard notation for partial derivatives: for a multi-index $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$, d being a positive integer, we denote by $|\alpha| = \alpha_1 + \cdots + \alpha_d$ its length and by ∂_α the differential operator $\partial^{|\alpha|}/\partial_1^{\alpha_1} \ldots \partial_d^{\alpha_d}$.

Lemma 32 — *Under assumptions ($\mathcal{H}12$), ($\mathcal{H}13$) and ($\mathcal{H}14$), we have that*

i) u_x is $\mathcal{C}^3$ with respect to the time variable and $\mathcal{C}^6$ with respect to the space variable. Moreover, it solves the following PDE

$$\left\{ \begin{array}{rcl} \partial_t u_x + \mathcal{L} u_x & = & 0 \\ u_x(T, y, m, v) & = & \gamma_x(y, m, v) \end{array} \right. \tag{2.24}$$

where $\mathcal{L}$ is the differential operator associated to (2.23):

$$\mathcal{L}u(y, m, v) = \frac{\sigma^2(y)}{2}\frac{\partial^2 u}{\partial y^2} + b(y)\frac{\partial u}{\partial y} + h(y)\frac{\partial u}{\partial m} + f^2(y)\frac{\partial u}{\partial v}.$$

ii) For any multi-index $\alpha \in \mathbb{N}^3$ and integer l such that $2l + |\alpha| \leq 6$, there exists $C_{l,\alpha}, K_{l,\alpha} > 0$ and $(p_{l,\alpha}, q_{l,\alpha}) \in \mathbb{N}^2$ such that

$$\forall (t,y,m,v) \in [0,T] \times \mathcal{D}_t, \quad \left|\partial_t^l \partial_\alpha u_x(t,y,m,v)\right| \leq C_{l,\alpha} e^{-K_{l,\alpha}x^2} \left(1+|x|^{p_{l,\alpha}}\right)\left(1+|y|^{q_{l,\alpha}}\right)$$

where $\mathcal{D}_t$ is the set $\mathbb{R} \times [-t\sup_{z\in\mathbb{R}}|h(z)|, t\sup_{z\in\mathbb{R}}|h(z)|] \times [t\underline{\psi}, t\overline{\psi}]$. Note that $\underline{\psi}$ and $\overline{\psi}$ are finite by virtue of assumptions ($\mathcal{H}13$) and ($\mathcal{H}14$).

Lemma 33 — *Under assumption ($\mathcal{H}12$),*

$$\forall q \in \mathbb{N}, \sup_{0\leq k\leq N} \mathbb{E}\left(\left|\overline{Y}^N_{t_k}\right|^q\right) < \infty$$

Now, following the error analysis of Talay and Tubaro [117], we write that

$$\left|\mathbb{E}\left[\gamma_x(Y_T, m_T, v_T) - \gamma_x(\overline{Y}^N_T, \overline{m}^N_T, \overline{v}^N_T)\right]\right| \leq \sum_{k=0}^{N-1} \eta_k(x)$$

where $\eta_k(x) = \left|\mathbb{E}\left[u_x(t_{k+1}, \overline{Y}^N_{t_{k+1}}, \overline{m}^N_{t_{k+1}}, \overline{v}^N_{t_{k+1}}) - u_x(t_k, \overline{Y}^N_{t_k}, \overline{m}^N_{t_k}, \overline{v}^N_{t_k})\right]\right|$ and $\forall 0 \leq k \leq N$, $\overline{m}^N_{t_k} = \delta_N \sum_{j=0}^{k-1} \frac{h(\overline{Y}^N_{t_j}) + h(\overline{Y}^N_{t_{j+1}})}{2}$ and $\overline{v}^N_{t_k} = \delta_N \sum_{j=0}^{k-1} \frac{f^2(\overline{Y}^N_{t_j}) + f^2(\overline{Y}^N_{t_{j+1}})}{2}$. Using the Markov property for the first term in the expectation and Taylor's formula together with PDE (2.24) for the second, we get

$$\begin{aligned}
\eta_k(x) \quad = \quad & \Big|\mathbb{E}\Big[\phi_x(t_{k+1}, \overline{Y}^N_{t_k}, \overline{m}^N_{t_k}, \overline{v}^N_{t_k}) - u_x(t_{k+1}, \overline{Y}^N_{t_k}, \overline{m}^N_{t_k}, \overline{v}^N_{t_k}) - \delta_N \mathcal{L} u_x(t_{k+1}, \overline{Y}^N_{t_k}, \overline{m}^N_{t_k}, \overline{v}^N_{t_k}) \\
& - \frac{\delta_N^2}{2}\mathcal{L}^2 u_x(t_{k+1}, \overline{Y}^N_{t_k}, \overline{m}^N_{t_k}, \overline{v}^N_{t_k}) + \frac{1}{2}\int_{t_k}^{t_{k+1}} \frac{\partial^3 u_x}{\partial t^3}(t, \overline{Y}^N_{t_k}, \overline{m}^N_{t_k}, \overline{v}^N_{t_k})(t-t_k)^2 dt\Big]\Big|
\end{aligned}$$

where

$$\phi_x(t_{k+1}, y, m, v) = \mathbb{E}\left[u_x(t_{k+1}, \overline{Y}^{N,y}_{t_1}, m + \delta_N \frac{h(\overline{Y}^{N,y}_{t_1}) + h(y)}{2}, v + \delta_N \frac{f^2(\overline{Y}^{N,y}_{t_1}) + f^2(y)}{2})\right]$$

Denote by Γ_y the function $z \mapsto u_x(t_{k+1}, z, m + \delta_N \frac{h(z)+h(y)}{2}, v + \delta_N \frac{f^2(z)+f^2(y)}{2})$. Using Taylor's formula we can show that $\forall z \in \mathbb{R}$,

$$\Gamma_y(z) = \Gamma_{y,1}(z) + \delta_N \Gamma_{y,2}(z) + \frac{\delta_N^2}{2}\Gamma_{y,3}(z) + R_0(z)$$

where

$$\begin{aligned}
\Gamma_{y,1}(z) &= u_x(t_{k+1}, z, m, v) \\
\Gamma_{y,2}(z) &= \frac{h(z)+h(y)}{2}\frac{\partial u_x}{\partial m}(t_{k+1}, z, m, v) + \frac{f^2(z)+f^2(y)}{2}\frac{\partial u_x}{\partial v}(t_{k+1}, z, m, v) \\
\Gamma_{y,3}(z) &= \left(\frac{h(z)+h(y)}{2}\right)^2 \frac{\partial^2 u_x}{\partial m^2}(t_{k+1}, z, m, v) + \left(\frac{f^2(z)+f^2(y)}{2}\right)^2 \frac{\partial^2 u_x}{\partial v^2}(t_{k+1}, z, m, v) \\
&\quad + 2\frac{h(z)+h(y)}{2}\frac{f^2(z)+f^2(y)}{2}\frac{\partial^2 u_x}{\partial m \partial v}(t_{k+1}, z, m, v)
\end{aligned}$$

and

$$
\begin{aligned}
R_0(z) = \int_0^{\delta_N} \frac{(\delta_N - t)^2}{2} dt \Bigg(& \left(\frac{h(z)+h(y)}{2}\right)^3 \frac{\partial^3 u_x}{\partial m^3}\left(t_{k+1}, z, m + t\frac{h(z)+h(y)}{2}, v + t\frac{f^2(z)+f^2(y)}{2}\right) \\
& + \left(\frac{f^2(z)+f^2(y)}{2}\right)^3 \frac{\partial^3 u_x}{\partial v^3}\left(t_{k+1}, z, m + t\frac{h(z)+h(y)}{2}, v + t\frac{f^2(z)+f^2(y)}{2}\right) \\
& +3\left(\frac{f^2(z)+f^2(y)}{2}\right)^2 \left(\frac{h(z)+h(y)}{2}\right) \frac{\partial^3 u_x}{\partial m \partial v^2}\left(t_{k+1}, z, m + t\frac{h(z)+h(y)}{2}, v + t\frac{f^2(z)+f^2(y)}{2}\right) \\
& +3\left(\frac{h(z)+h(y)}{2}\right)^2 \left(\frac{f^2(z)+f^2(y)}{2}\right) \frac{\partial^3 u_x}{\partial m^2 \partial v}\left(t_{k+1}, z, m + t\frac{h(z)+h(y)}{2}, v + t\frac{f^2(z)+f^2(y)}{2}\right)\Bigg)
\end{aligned}
\tag{2.25}
$$

So,

$$
\phi_x(t_{k+1}, y, m, v) = \underbrace{\mathbb{E}\left[\Gamma_{y,1}(\overline{Y}_{t_1}^{N,y})\right]}_{\phi_{x,1}(t_{k+1},y,m,v)} + \underbrace{\delta_N \mathbb{E}\left[\Gamma_{y,2}(\overline{Y}_{t_1}^{N,y})\right]}_{\phi_{x,2}(t_{k+1},y,m,v)} + \underbrace{\frac{\delta_N^2}{2}\mathbb{E}\left[\Gamma_{y,3}(\overline{Y}_{t_1}^{N,y})\right]}_{\phi_{x,3}(t_{k+1},y,m,v)} + \mathbb{E}\left[R_0(\overline{Y}_{t_1}^{N,y})\right] \tag{2.26}
$$

With a slight abuse of notations, we define the first order differential operators V_0 and V acting on $\mathcal{C}^1$ functions by $V_0\xi(x) = V_0(x)\xi'(x)$ and $V\xi(x) = V(x)\xi'(x)$ for $\xi \in \mathcal{C}^1(\mathbb{R})$. We make the same expansions as in Ninomiya and Victoir [98] but with making the remainder terms explicit in order to check if they have the good behavior with respect to x. We can show after tedious but simple computations that

$$
\begin{aligned}
\phi_{x,1}(t_{k+1}, y, m, v) &= \Gamma_{y,1}(y) + \frac{\delta_N}{2}\left(V^2\Gamma_{y,1}(y) + 2V_0\Gamma_{y,1}(y)\right) \\
&\quad + \frac{\delta_N^2}{8}\left({4V_0}^2\Gamma_{y,1}(y) + 2V_0V^2\Gamma_{y,1}(y) + 2V^2V_0\Gamma_{y,1}(y) + V^4\Gamma_{y,1}(y)\right) + \mathbb{E}\left(R_1(y)\right) \\
\phi_{x,2}(t_{k+1}, y, m, v) &= \delta_N\Gamma_{y,2}(y) + \frac{\delta_N^2}{2}\left(V^2\Gamma_{y,2}(y) + 2V_0\Gamma_{y,2}(y)\right) + \mathbb{E}\left(R_2(y)\right) \\
\phi_{x,3}(t_{k+1}, y, m, v) &= \frac{\delta_N^2}{2}\Gamma_{y,3}(y) + \mathbb{E}\left(R_3(y)\right)
\end{aligned}
$$

where

$$
\begin{aligned}
R_1(y) &= \int_0^{\frac{\delta_N}{2}} \int_0^{s_1} \int_0^{s_2} {V_0}^3 \Gamma_{y,1}(e^{s_3 V_0} e^{W_{\delta_N} V} e^{\frac{\delta_N}{2} V_0}(y)) ds_3 ds_2 ds_1 \\
&+ \int_0^{W_{\delta_N}} \int_0^{s_1} \int_0^{s_2} \int_0^{s_3} \int_0^{s_4} \int_0^{s_5} V^6 \Gamma_{y,1}(e^{s_6 V} e^{\frac{\delta_N}{2} V_0}(y)) ds_6 ds_5 ds_4 ds_3 ds_2 ds_1 \\
&+ \frac{\delta_N}{2} \int_0^{W_{\delta_N}} \int_0^{s_1} \int_0^{s_2} \int_0^{s_3} V^4 V_0 \Gamma_{y,1}(e^{s_4 V} e^{\frac{\delta_N}{2} V_0}(y)) ds_4 ds_3 ds_2 ds_1 \\
&+ \frac{\delta_N^2}{8} \int_0^{W_{\delta_N}} \int_0^{s_1} V^2 {V_0}^2 \Gamma_{y,1}(e^{s_2 V} e^{\frac{\delta_N}{2} V_0}(y)) ds_2 ds_1 \\
&+ \int_0^{\frac{\delta_N}{2}} \int_0^{s_1} \int_0^{s_2} {V_0}^3 \Gamma_{y,1}(e^{s_3 V_0}(y)) ds_3 ds_2 ds_1 + \frac{\delta_N}{2} \int_0^{\frac{\delta_N}{2}} \int_0^{s_1} {V_0}^2 V^2 \Gamma_{y,1}(e^{s_2 V_0}(y)) ds_2 ds_1 \\
&+ \frac{\delta_N^2}{8} \int_0^{\frac{\delta_N}{2}} V_0 V^4 \Gamma_{y,1}(e^{s_1 V_0}(y)) ds_1 + \frac{\delta_N}{2} \int_0^{\frac{\delta_N}{2}} \int_0^{s_1} {V_0}^3 \Gamma_{y,1}(e^{s_2 V_0}(y)) ds_2 ds_1 \\
&+ \frac{\delta_N^2}{4} \int_0^{\frac{\delta_N}{2}} V_0 V^2 V_0 \Gamma_{y,1}(e^{s_1 V_0}(y)) ds_1 + \frac{\delta_N^2}{8} \int_0^{\frac{\delta_N}{2}} {V_0}^3 \Gamma_{y,1}(e^{s_1 V_0}(y)) ds_1 \\
R_2(y) &= \delta_N \Bigg(\int_0^{\frac{\delta_N}{2}} \int_0^{s_1} {V_0}^2 \Gamma_{y,2}(e^{s_2 V_0} e^{W_{\delta_N} V} e^{\frac{\delta_N}{2} V_0}(y)) ds_2 ds_1 \\
&+ \int_0^{W_{\delta_N}} \int_0^{s_1} \int_0^{s_2} \int_0^{s_3} V^4 \Gamma_{y,2}(e^{s_4 V} e^{\frac{\delta_N}{2} V_0}(y)) ds_4 ds_3 ds_2 ds_1 \\
&+ \frac{\delta_N}{2} \int_0^{W_{\delta_N}} \int_0^{s_1} V^2 V_0 \Gamma_{y,2}(e^{s_2 V} e^{\frac{\delta_N}{2} V_0}(y)) ds_2 ds_1 + \int_0^{\frac{\delta_N}{2}} \int_0^{s_1} {V_0}^2 \Gamma_{y,2}(e^{s_2 V_0}(y)) ds_2 ds_1 \\
&+ \frac{\delta_N}{2} \int_0^{\frac{\delta_N}{2}} V_0 V^2 \Gamma_{y,2}(e^{s_1 V_0}(y)) ds_1 + \frac{\delta_N}{2} \int_0^{\frac{\delta_N}{2}} {V_0}^2 \Gamma_{y,2}(e^{s_1 V_0}(y)) ds_1 \Bigg) \\
R_3(y) &= \frac{\delta_N^2}{2} \Bigg(\int_0^{\frac{\delta_N}{2}} V_0 \Gamma_{y,3}(e^{s_1 V_0} e^{W_{\delta_N} V} e^{\frac{\delta_N}{2} V_0}(y)) ds_1 + \int_0^{W_{\delta_N}} \int_0^{s_1} V^2 \Gamma_{y,3}(e^{s_2 V} e^{\frac{\delta_N}{2} V_0}(y)) ds_2 ds_1 \\
&+ \int_0^{\frac{\delta_N}{2}} V_0 \Gamma_{y,3}(e^{s_1 V_0}(y)) ds_1 \Bigg)
\end{aligned}
\tag{2.27}
$$

Putting all the terms together, one can check that

$$
\phi_x(t_{k+1}, y, m, v) = u_x(t_{k+1}, y, m, v) + \delta_N \mathcal{L} u_x(t_{k+1}, y, m, v) + \frac{\delta_N^2}{2} \mathcal{L}^2 u_x(t_{k+1}, y, m, v) + R(y)
$$

where $R(y) = \mathbb{E}\left[R_0(\overline{Y}_{t_1}^{N,y}) + R_1(y) + R_2(y) + R_3(y) \right]$. Finally,

$$
\left| \mathbb{E}\left[\gamma_x(Y_T, m_T, v_T) - \gamma_x(\overline{Y}_T^N, \overline{m}_T^N, \overline{v}_T^N) \right] \right| \leq \sum_{k=0}^{N-1} \mathbb{E}\left[\left| \frac{1}{2} \int_{t_k}^{t_{k+1}} \frac{\partial^3 u_x}{\partial t^3}(t, \overline{Y}_{t_k}^N, \overline{m}_{t_k}^N, \overline{v}_{t_k}^N)(t - t_k)^2 dt \right| + \left| R(\overline{Y}_{t_k}^N) \right| \right]
$$

From Lemmas 32 and 33, we deduce that there exists $C_1, K_1 > 0$ and $p_1 \in \mathbb{N}$ such that

$$
\sum_{k=0}^{N-1} \mathbb{E}\left[\left| \frac{1}{2} \int_{t_k}^{t_{k+1}} \frac{\partial^3 u_x}{\partial t^3}(t, \overline{Y}_{t_k}^N, \overline{m}_{t_k}^N, \overline{v}_{t_k}^N)(t - t_k)^2 dt \right| \right] \leq \frac{1}{N^2} C_1 e^{-K_1 x^2} (1 + |x|^{p_1}) \tag{2.28}
$$

On the other hand, a close look to (2.25) and (2.27) convinces us that the term $\mathbb{E}\left[\left| R(\overline{Y}_{t_k}^N) \right| \right]$ is of order $\frac{1}{N^3}$ and that it involves only derivatives of u_x and of the coefficients of the SDE (2.23).

So, thanks Lemmas 32 and 33, there exists $C_2, K_2 > 0$ and $p_2 \in \mathbb{N}$ such that

$$\sum_{k=0}^{N-1} \mathbb{E}\left[\left|R(\overline{Y}^N_{t_k})\right|\right] \leq \frac{1}{N^2} C_2 e^{-K_2 x^2}(1+|x|^{p_2}) \tag{2.29}$$

From (2.28) and (2.29) we deduce the desired result (2.22) to conclude.

□

Remark 34 —

- *The theorem does not cover the case of perfectly correlated or uncorrelated stock and volatility which is not very interesting from a practical point of view.*
- *As for plain vanilla options pricing, observe that, by the Romano and Touzi [106] formula,*

$$\mathbb{E}\left(e^{-rT}\alpha(S_T)|(Y_t)_{t\in[0,T]}\right) = BS_{\alpha,T}\left(s_0 e^{\rho(F(Y_T)-F(y_0))+m_T+(\frac{(1-\rho^2)v_T}{2T}-r)T}, \frac{(1-\rho^2)v_T}{T}\right)$$

where $BS_{\alpha,T}(s,v)$ stands for the price of a European option with pay-off α and maturity T in the Black & Scholes model with initial stock price s, volatility $\sqrt{v}$ and constant interest rate r. When, like for a call or a put option, $BS_{\alpha,T}$ is available in a closed form, one should approximate $\mathbb{E}\left(e^{-rT}\alpha(S_T)\right)$ by

$$\frac{1}{M}\sum_{i=1}^{M} BS_{\alpha,T}\left(s_0 e^{\rho(F(\overline{Y}^{N,i}_T)-F(y_0))+\overline{m}^{N,i}_T+(\frac{(1-\rho^2)\overline{v}^{N,i}_T}{2T}-r)T}, \frac{(1-\rho^2)\overline{v}^{N,i}_T}{T}\right)$$

where M is the total number of Monte Carlo samples and the index i refers to independent draws.

Indeed, the conditioning provides a variance reduction. We also note that what is most important is to have a scheme with a high order weak convergence on the triplet $(Y_t, m_t, v_t)_{t\in[0,T]}$ solution of the SDE (2.23), which is the case for our scheme.

- *In the special case of an Ornstein-Uhlenbeck process driving the volatility (i.e $(Y_t)_{t\in[0,T]}$ is solution of the SDE (2.11)), one should replace the Ninomiya-Victoir scheme by the true solution. We can then prove more easily the same weak convergence result: at step (2.26) of the preceding proof, we apply Itô's formula instead of carrying out the Ninomiya-Victoir expansion. Moreover, we can prove, following the same error analysis, that the OU_Improved scheme (2.14) also exhibits a second order weak convergence property. Better still, it achieves a weak trajectorial convergence of order $\frac{3}{2}$ on the triplet $(Y_t, m_t, v_t)_{t\in[0,T]}$ which allows for a significant improvement of the multilevel Monte Carlo method, as we shall check numerically.*

2.3 Numerical results

For numerical computations, we are going to consider Scott's model (2.2). We use the same set of parameters as in Kahl and Jäckel [61] : $S_0 = 100, r = 0.05, T = 1, y_0 = \log(0.25), \kappa = 1, \theta = 0, \nu = \frac{7\sqrt{2}}{20}, \rho = -0.2$ and $f : y \mapsto e^y$.

We are going to compare our schemes (WeakTraj_1, Weak_2 and OU_Improved) to the Euler scheme with exact simulation of the volatility (hereafter denoted Euler), the Kahl and Jäckel [61] scheme (IJK) and the Cruzeiro *et al.* [27] scheme (CMT).

2.3.1 Numerical illustration of strong convergence properties

In order to illustrate the strong convergence rate of a discretization scheme $\widehat{X}^N$, we consider the squared L^2-norm of the supremum of the difference between the scheme with time step $\frac{T}{N}$ and the one with time step $\frac{T}{2N}$:

$$\mathbb{E}\left[\max_{0\leq k\leq N}\left|\widehat{X}^N_{t_k}-\widehat{X}^{2N}_{t_k}\right|^2\right] \tag{2.30}$$

This quantity will exhibit the same asymptotic behavior with respect to N as the squared L^2-norm of the difference between the scheme with time step $\frac{T}{N}$ and the limiting process towards which it converges (see Alfonsi [1]).

In Figure 2.1, we draw the logarithm of the Monte Carlo estimation of (2.30) as a function of the logarithm of the number of time steps. The number of Monte Carlo samples used is equal to $M = 10000$ and the number of discretization steps is a power of 2 varying from 2 to 256. We also consider the strong convergence of the schemes on the asset itself (see Figure 2.2) by computing $\mathbb{E}\left[\max_{0\leq k\leq N}\left|e^{\widehat{X}^N_{t_k}}-e^{\widehat{X}^{2N}_{t_k}}\right|^2\right]$.

The slopes of the regression lines are reported in Table 2.1. We see that, both for the logarithm of the asset and for the asset itself, all the schemes exhibit a strong convergence of order $\frac{1}{2}$. Our schemes only have a better constant.

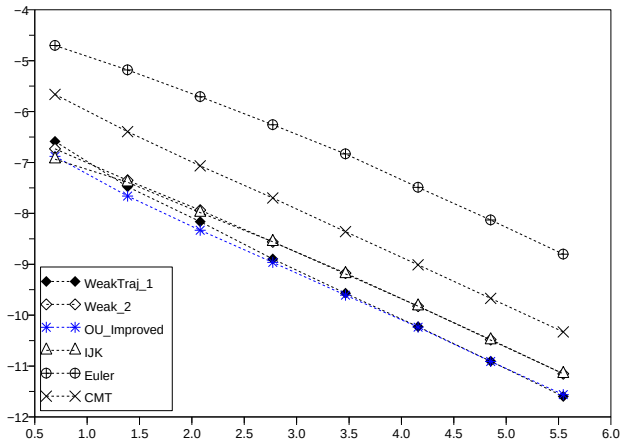

Figure 2.1: Strong convergence on the log-asset

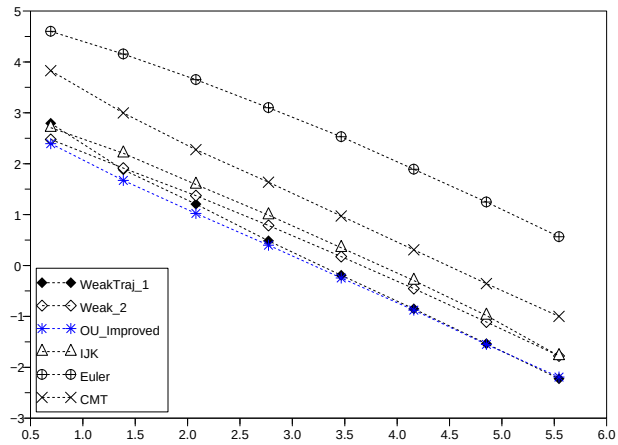

Figure 2.2: Strong convergence on the asset

	WeakTraj_1	Weak_2	OU_Improved	IJK	CMT	Euler
Log-asset	-1.01	-0.88	-0.94	-0.92	-0.98	-0.84
Asset	-1.01	-0.91	-0.95	-0.88	-0.95	-0.85

Table 2.1: Slopes of the regression lines (Strong convergence)

Weak trajectorial convergence

Nevertheless, as explained in Remark 23, for the scheme with time step $\frac{1}{N}$, one can replace the increments of the Brownian motion $(B_t)_{t\in[0,T]}$ by a sequence of Gaussian random variables smartly constructed from the scheme with time step $\frac{1}{2N}$. This particular coupling is possible whenever the independence structure between $(B_t)_{t\in[0,T]}$ and $(Y_t)_{t\in[0,T]}$ is preserved by the discretization of the latter process, which is the case for all the schemes but the CMT scheme. So we carry out this coupling and we repeat the preceding numerical experiment. The results are put together in Figures 2.3 and 2.4 and in Table 2.2.

As expected, we see that the WeakTraj_1 and the OU_Improved schemes exhibit a first order convergence rate whereas the other schemes exhibit a $\frac{1}{2}$ order convergence rate. Note that the CMT scheme has a weak trajectorial convergence of order one but it is much more difficult to implement the coupling for which the convergence order is indeed equal to one.

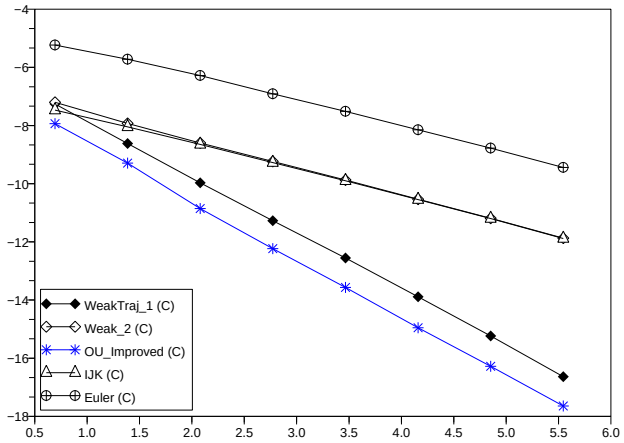

Figure 2.3: Weak trajectorial convergence on the log-asset (with coupling)

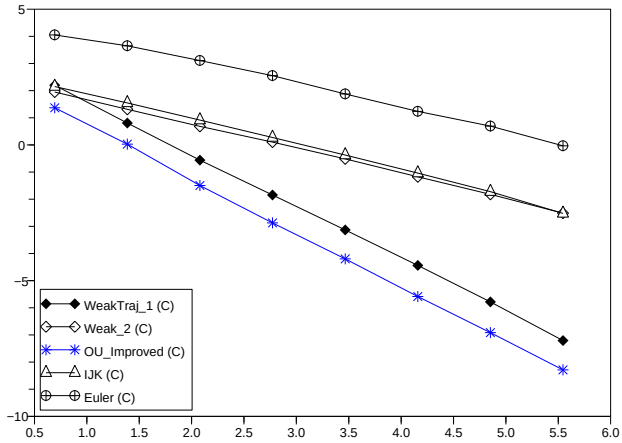

Figure 2.4: Weak trajectorial convergence on the asset (with coupling)

	WeakTraj_1	Weak_2	OU_Improved	IJK	CMT	Euler
Log-asset	-1.92	-0.91	-1.99	-0.95	–	-0.85
Asset	-1.92	-0.95	-2	-0.91	–	-0.87

Table 2.2: Slopes of the regression lines (Weak trajectorial convergence)

Convergence at terminal time

We consider now convergence at terminal time, precisely the squared L^2-norm of the difference between the terminal values of the schemes with time steps $\frac{T}{N}$ and $\frac{T}{2N}$:

$$\mathbb{E}\left[\left|\widehat{X}_T^N - \widehat{X}_T^{2N}\right|^2\right]. \tag{2.31}$$

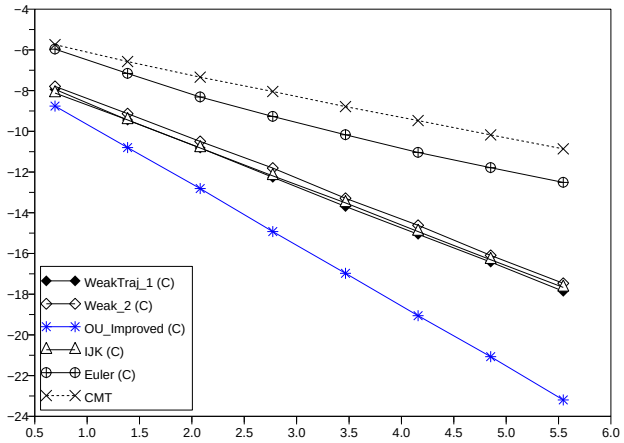

Figure 2.5: Convergence at terminal time for the log-asset

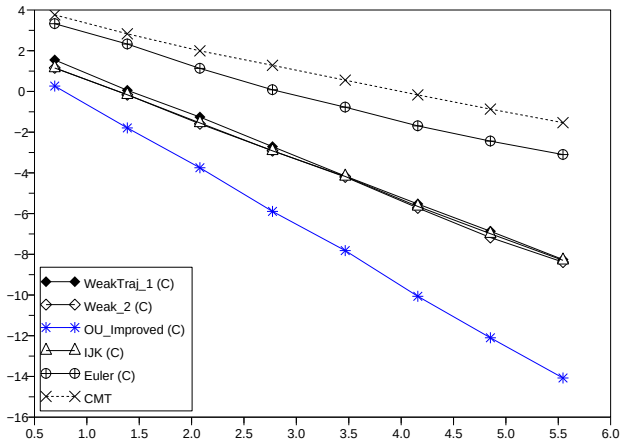

Figure 2.6: Convergence at terminal time for the asset

	WeakTraj_1	Weak_2	OU_Improved	IJK	CMT	Euler
Log-asset	-2.03	-2	-2.97	-1.97	-1.05	-1.34
Asset	-2.02	-1.98	-2.97	-1.95	-1.08	-1.34

Table 2.3: Slopes of the regression lines (Convergence at terminal time)

Note that we introduce a coupling : we write the schemes straight at the terminal time as we did for the Weak_2 scheme (see (2.20)) and we generate the terminal values of the schemes with time steps $\frac{T}{N}$ and $\frac{T}{2N}$ using the same single normal random variable to simulate the stochastic integral w.r.t. $(B_t)_{t\in[0,T]}$. Once again, it is possible to proceed alike for all the schemes but the CMT scheme. For the latter, we simulate the scheme at all the intermediate discretization times to obtain the value at terminal time.

We also consider the convergence at terminal time of the asset itself. We report the numerical results in Figures 2.5 and 2.6 and give the slopes of the regression lines in Table 2.3.

We observe that, as stated in Remark 28, the OU_Improved scheme exhibits a convergence rate of order $\frac{3}{2}$, outperforming all the other schemes. As previously, the WeakTrak_1 scheme exhibits a first order convergence rate. Note also that this new coupling at terminal time improved the convergence rate of the Weak_2 and the IJK schemes up to order one and, surprisingly, it improved the convergence rate of the Euler scheme up to an order strictly greater than the expected $\frac{1}{2}$, approximately 0.67.

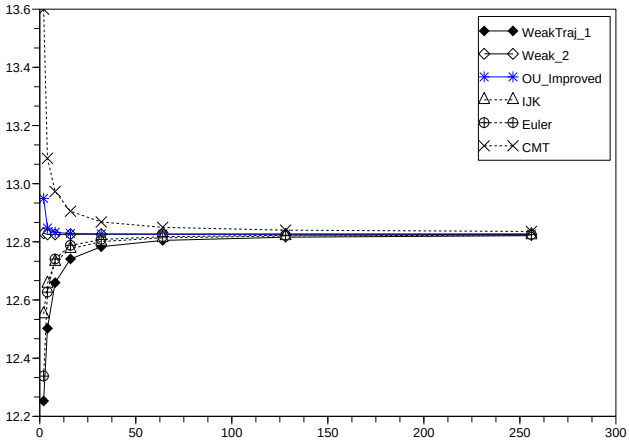

Figure 2.7: Convergence of the call price with respect to N

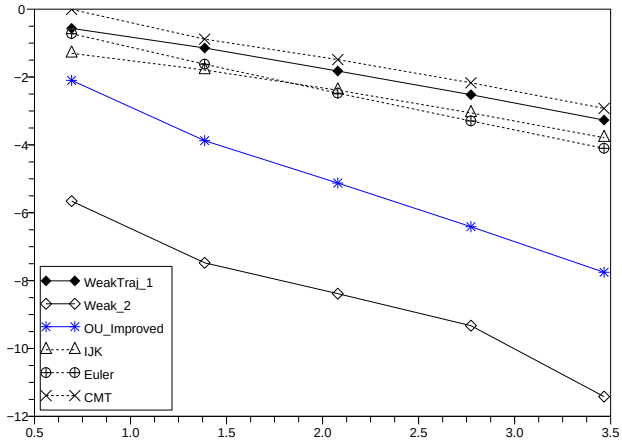

Figure 2.8: Illustration of the convergence rate for the call option

2.3.2 Standard call pricing

Numerical illustration of weak convergence

We compute the price of a call option with strike $K = 100$ and maturity $T = 1$. For all the schemes but the CMT scheme, we use the conditioning variance reduction technique presented in Remark 34.

In Figure 2.7, we draw the price as a function of the number of time steps for each scheme and in Figure 2.8 we draw the logarithm of the pricing error : $\log\left(\left|P_{\text{exact}} - P^N_{\text{scheme}}\right|\right)$ where $P_{\text{exact}} \approx 12.82603$ is obtained by a multilevel Monte Carlo with an accuracy of $5bp$, as a function of the logarithm of the number of times steps.

We see that, as expected, the Weak_2 scheme and the OU_Improved scheme exhibit a weak convergence of order two and converge much faster than the others. The weak scheme already gives an accurate price with only four time steps. The WeakTraj_1 scheme has a weak convergence of order one like the Euler and the IJK scheme, but it has a greater leading error term. Fortunately, its better strong convergence properties enable it to catch up with the multilevel Monte Carlo method as we will see hereafter.

Finally, note that the weak scheme does not require the simulation of additional terms when compared to the Euler or the IJK schemes. Combined with its second order weak convergence order, this makes the Weak_2 scheme very competitive for the pricing of plain vanilla European option.

Multilevel monte carlo

Let us now apply the multilevel Monte Carlo method of Giles [48] to compute the Call price. As previously, we consider the schemes straight at the terminal time and use a conditioning variance reduction technique. We give the CPU time as a function of the root mean square error in Figure

2.9 (see Giles [48] for details on the heuristic numerical algorithm which is used).

We observe that both the Weak_2 and the OU_Improved scheme are great time-savers. For the OU_Improved scheme, the effect coming from its good strong convergence properties is somewhat offset by the additional terms it requires to simulate. We can see nevertheless that it is going to overcome the Weak_2 scheme for bigger accuracy levels.

2.3.3 Asian option pricing and multilevel Monte Carlo

Finally, we consider an example of path-dependent option pricing : the Asian option. More precisely, we compute the price of the Asian call option with strike $K = 100$ whose pay-off is equal to $\left(\frac{1}{T}\int_0^T S_t dt - K\right)_+$ and we choose to discretize the integral of the stock price by a trapezoidal method for each scheme.

We first draw the price obtained by the different schemes with respect to the number of time steps N (see Figure 2.10) and the logarithm of the pricing error : $\log\left(\left|P_{\text{exact}} - P^N_{\text{scheme}}\right|\right)$ where $P_{\text{exact}} \approx 7.0364$ is obtained by a multilevel Monte Carlo with an accuracy of $5bp$, as a function of the logarithm of the number of times steps (see Figure 2.11). For all the schemes but the OU_Improved scheme, the convergence rates seem to be quite similar, around one. Surprisingly, the OU_Improved scheme exhibits a second order convergence and far outperforms all the other schemes. For example, it achieves the same precision for $N = 16$ as the other schemes for $N = 128$. The WeakTraj_1 scheme is a little bit slower than the Weak_2, the IJK and the Euler schemes.

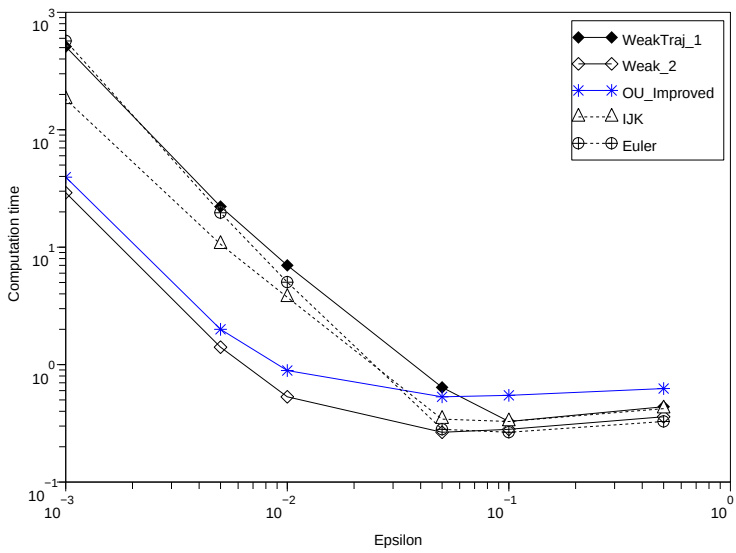

Figure 2.9: Multilevel Monte Carlo method for a Call option using different schemes

However, as explained in Remark 23, the main advantage of this scheme is that it improves the convergence of the multilevel Monte Carlo method. In Figure 2.12, we draw the CPU time times the mean square error against the root mean square error.

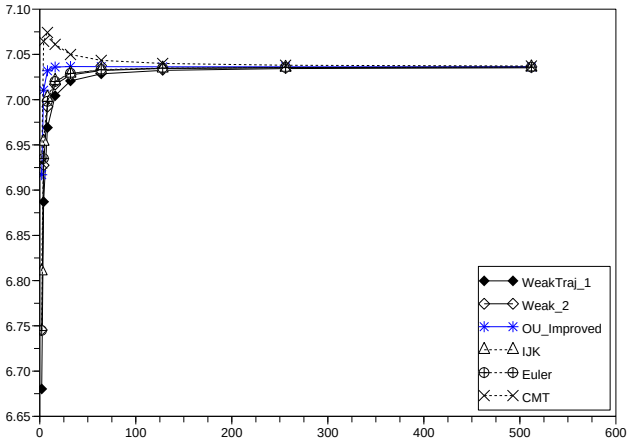

Figure 2.10: Convergence of the Asian price with respect to N

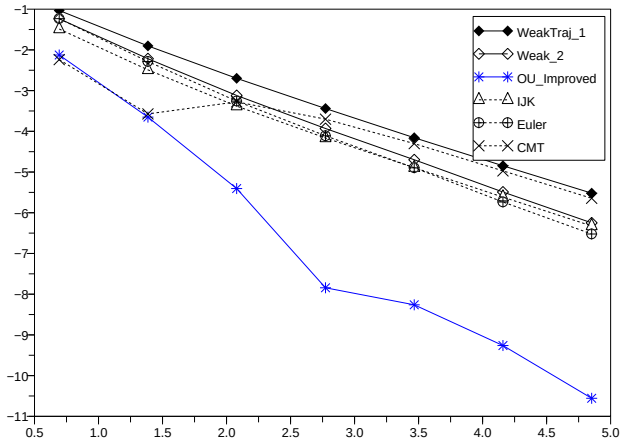

Figure 2.11: Illustration of the convergence rate for the Asian option

We see that our schemes perform better than the others. Certainly, the gain obtained is not as important as for the call pricing example. This is maybe due to the fact that the good strong convergence properties of our schemes are hidden by the discretization bias coming from the approximation of the integral in time of the asset price with a finite sum.

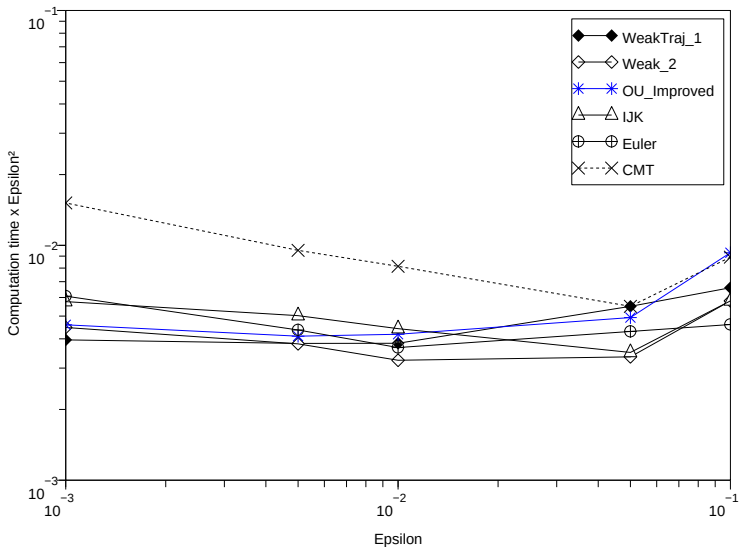

Figure 2.12: Multilevel Monte Carlo method for an Asian option using different schemes.

2.4 Conclusion

In this article, we have capitalized on the particular structure of stochastic volatility models to propose and discuss two simple and yet competitive discretization schemes. The first one exhibits first order weak trajectorial convergence and has the advantage of improving multilevel Monte Carlo methods for the pricing of path dependent options. The second one is rather useful for pricing European options since it has a second order weak convergence rate.

We have also focused on the special case of an Ornstein-Uhlenbeck process driving the volatility, which encompasses many stochastic volatility models such as the Scott [112]'s model or the quadratic Gaussian model. Then, the convergence properties of the previous schemes are preserved when simulating $(Y_t)_{0\leq t\leq T}$ exactly. We have also proposed an improved scheme exhibiting both weak trajectorial convergence of order one and weak convergence of order two.

The numerical experiments show that our schemes are very competitive for the pricing of plain vanilla and path-dependent options. Their use with multilevel Monte Carlo gives satisfactory results too. We should also mention that the main purpose of our study was the convergence order with respect to the time step. It would be of great interest to carry out an extensive numerical study of the computational complexity of the schemes presented in this paper. This will be the subject of future research.

2.5 Appendix

2.5.1 Proof of Lemma 21

We first suppose that $p = 1$. According to Theorem 5.2 page 72 of Milstein [92], it suffices to check that there exists a positive constant C independent of N such that

$$\begin{aligned} \left|\mathbb{E}\left(Y_{\delta_N} - \overline{Y}^N_{\delta_N}\right)\right| &\leq C\delta_N^2 \\ \left|\mathbb{E}\left(\left(Y_{\delta_N} - \overline{Y}^N_{\delta_N}\right)^2\right)\right|^{\frac{1}{2}} &\leq C\delta_N^{\frac{3}{2}} \\ \left|\mathbb{E}\left(\left(Y_{\delta_N} - \overline{Y}^N_{\delta_N}\right)^4\right)\right|^{\frac{1}{4}} &\leq C\delta_N^{\frac{5}{4}} \end{aligned} \tag{2.32}$$

First note that

$$Y_{\delta_N} - \overline{Y}^N_{\delta_N} = \int_0^{\delta_N} b(Y_s) - b(y_0)ds + \int_0^{\delta_N}\left(\int_0^s (b\sigma' + \frac{1}{2}\sigma^2\sigma'')(Y_r)dr + (\sigma\sigma'(Y_r) - \sigma\sigma'(y_0))dW_r\right)dW_s$$

Thanks to Itô's formula and to assumption $(\mathcal{H}5)$, we have that

$$\begin{aligned} \left|\mathbb{E}\left(Y_{\delta_N} - \overline{Y}^N_{\delta_N}\right)\right| &= \left|\int_0^{\delta_N}\int_0^s \mathbb{E}\left((bb' + \frac{1}{2}b''\sigma^2)(Y_r)\right)drds\right| \\ &\leq C\left|\int_0^{\delta_N}\int_0^s C(1 + \mathbb{E}(|Y_r|^2))drds\right| \\ &\leq C\delta_N^2 \end{aligned}$$

Using assumptions ($\mathcal{H}5$) and ($\mathcal{H}6$), we also have $\forall p \geq 1$

$$\begin{aligned}
\mathbb{E}\left(\left|Y_{\delta_N} - \overline{Y}^N_{\delta_N}\right|^{2p}\right) &\leq 2^{2p-1}\mathbb{E}\left[\left|\int_0^{\delta_N} b(Y_s) - b(y_0)ds\right|^{2p}\right. \\
&\quad \left. + \left|\int_0^{\delta_N}\left(\int_0^s (b\sigma' + \frac{1}{2}\sigma\sigma'')(Y_r)dr + (\sigma\sigma'(Y_r) - \sigma\sigma'(y_0))dW_r\right)dW_s\right|^{2p}\right] \\
&\leq 2^{2p-1}\left[\delta_N^{2p-1}\int_0^{\delta_N}\mathbb{E}\left(|b(Y_s) - b(y_0)|^{2p}\right)ds\right. \\
&\quad \left. + C\delta_N^{p-1}\int_0^{\delta_N}\mathbb{E}\left(\left|\int_0^s (b\sigma' + \frac{1}{2}\sigma\sigma'')(Y_r)dr + (\sigma\sigma'(Y_r) - \sigma\sigma'(y_0))dW_r\right|^{2p}\right)ds\right] \\
&\leq C\left[\delta_N^{2p-1}\int_0^{\delta_N} s^p ds + \delta_N^{p-1}\int_0^{\delta_N} s^{2p-1}\int_0^s \mathbb{E}\left(\left|(b\sigma' + \frac{1}{2}\sigma\sigma'')(Y_r)\right|^{2p}\right)dr\,ds\right. \\
&\quad \left. + \delta_N^{p-1}\int_0^{\delta_N} s^{p-1}\int_0^s \mathbb{E}\left(\left|\sigma\sigma'(Y_r) - \sigma\sigma'(y_0)\right|^{2p}\right)dr\,ds\right] \\
&\leq C\delta_N^{3p}
\end{aligned}$$

This implies both the second and the third inequality of (2.32). This estimation is also sufficient to extend the result of Milstein [92] to the L^{2p} norm and conclude the proof.

2.5.2 Proof of Lemma 26

One can easily check that $(Y_t)_{0\leq t\leq T}$ is a Gaussian process which has the same distribution law as the process $(y_0e^{-\kappa t} + \theta(1-e^{-\kappa t}) + \frac{\nu e^{-\kappa t}}{\sqrt{2\kappa}}W_{e^{2\kappa t}-1})_{0\leq t\leq T}$. So,

$$\begin{aligned}
\mathbb{E}\left(e^{c_1\sup_{0\leq t\leq T}|Y_t|^{1+c_2}}\right) &= \mathbb{E}\left(e^{c_1\sup_{0\leq t\leq T}|y_0e^{-\kappa t}+\theta(1-e^{-\kappa t})+\frac{\nu e^{-\kappa t}}{\sqrt{2\kappa}}W_{e^{2\kappa t}-1}|^{1+c_2}}\right) \\
&\leq C\mathbb{E}\left(e^{C\sup_{0\leq t\leq T}|W_{e^{2\kappa t}-1}|^{1+c_2}}\right)
\end{aligned}$$

Since $\sup_{0\leq t\leq e^{2\kappa T}-1}|W_t| = \left(\sup_{0\leq t\leq e^{2\kappa T}-1}W_t\right) \vee \left(-\inf_{0\leq t\leq e^{2\kappa T}-1}W_t\right)$, we deduce from the symmetry property of the Brownian motion that

$$\begin{aligned}
\mathbb{E}\left(e^{c_1\sup_{0\leq t\leq T}|Y_t|^{1+c_2}}\right) &\leq C\mathbb{E}\left(e^{C|\sup_{0\leq t\leq e^{2\kappa T}-1}W_t|^{1+c_2}} + e^{C|\inf_{0\leq t\leq e^{2\kappa T}-1}W_t|^{1+c_2}}\right) \\
&\leq 2C\mathbb{E}\left(e^{C|\sup_{0\leq t\leq e^{2\kappa T}-1}W_t|^{1+c_2}}\right)
\end{aligned}$$

The probability density function of $\sup_{0\leq t\leq T}W_t$ is equal to $y \mapsto \sqrt{\frac{2}{\pi T}}e^{-\frac{y^2}{2T}}\mathbb{1}_{\{y>0\}}$ (see for example problem 8.2 p. 96 of Karatzas and Shreve [63]) which permits to conclude.

2.5.3 Proof of Lemma 32

The first point is an obvious consequence of the Feynman-Kac theorem. In order to prove the second one, let us first check the following result :

$$\begin{aligned}
&\text{For any multi-index } \beta \in \mathbb{N}^3 \text{ such that } \beta_1 \leq 6, \exists\, C_\beta, K_\beta \geq 0 \text{ and } p_\beta \in \mathbb{N} \text{ such that} \\
&\forall (y,m,v) \in \mathcal{D}_T, \quad |\partial_\beta\gamma_x(y,m,v)| \leq C_\beta e^{-K_\beta x^2}\left(1 + |x|^{p_\beta}\right)
\end{aligned} \tag{2.33}$$

Indeed, using Leibniz's formula, one can show that $\partial_\beta \gamma_x(y,m,v)$ can be written as a weighted sum of terms of the form

$$\zeta_k = \frac{(x-\log(s_0)+\rho F(y_0)-\rho F(y)-m)^{k_2}}{v^{k_1+\frac{1}{2}}} \exp\left(-\frac{(x-\log(s_0)+\rho F(y_0)-\rho F(y)-m)^2}{2(1-\rho^2)v}\right)\prod_{i=0}^{k_3} a_i F^{(i)}(y)$$

where $k=(k_1,k_2,k_3)$ belongs to a finit set $I_\beta \subset \mathbb{N}^3$ and $(a_i)_{0\le i\le k_3}$ are constants taking value in $\{0,1\}$. Using assumption $(\mathcal{H}13)$ and $(\mathcal{H}14)$ and Young's inequality, we show that $\exists C_k, K_k > 0$ and $p_k \in \mathbb{N}$ such that $|\zeta_k| \le C_k e^{-K_k x^2}(1+|x|^{p_k})$ which yields the desired result.

Now, let us fix $\alpha \in \mathbb{N}^3, l \in \mathbb{N}$ such that $2l+|\alpha| \le 6$ and $(t,y,m,v) \in [0,T]\times \mathcal{D}_t$. Thanks to PDE (2.23), $\partial_t^l \partial_\alpha u_x(t,y,m,v) = (-1)^l \partial_\alpha \mathcal{L}^l u_x(t,y,m,v)$. One can check that the right hand side is equal to a weighted sum of terms of the form $\partial_{\beta_1} u_x(t,y,m,v) \times \pi_{\beta_2}(b,\sigma,f,h)$ where $\beta_1 \in \mathbb{N}^3$ is multi-index belonging to a finite set $I^1_{\alpha,l}$, β_2 is a suffix belonging to a finite set $I^2_{\alpha,l}$ and $\pi_{\beta_2}(b,\sigma,f,h)$ is a product of terms involving the functions b,σ,f,h and their derivatives up to order 4.

On the first hand, assumptions $(\mathcal{H}12)$ and $(\mathcal{H}13)$ yield that $\exists c^2_{l,\alpha} \ge 0$ and $q_{l,\alpha} \in \mathbb{N}$ such that

$$\forall \beta_2 \in I^2_{\alpha,l}, |\pi_{\beta_2}(b,\sigma,f,h)| \le c^2_{l,\alpha}(1+|y|^{q_{l,\alpha}}). \tag{2.34}$$

On the other hand, by inverting expectation and differentiations, we see that $\partial_{\beta_1} u_x(t,y,m,v)$ is equal to the expectation of a product between derivatives of the flow $(y,m,v) \to (Y_{T-t}, m_{T-t}, v_{T-t})^{(y,m,v)}$ and derivatives of the function γ_x evaluated at $(Y_{T-t}, m_{T-t}, v_{T-t})^{(y,m,v)} \in \mathcal{D}_T$. Using result (2.33) and the fact that, under assumptions $(\mathcal{H}12)$ and $(\mathcal{H}13)$, the derivatives of the flow satisfy a system of SDEs with Lipschitz continuous coefficients (see for example Kunita [74]) we show that $\exists c^1_{l,\alpha}, K_{l,\alpha} > 0$ and $p_{l,\alpha} \in \mathbb{N}$ such that

$$\forall \beta_1 \in I^1_{\alpha,l}, |\partial_{\beta_1} u_x(t,y,m,v)| \le c^1_{l,\alpha} e^{-K_{l,\alpha}x^2}(1+|x|^{p_{l,\alpha}}). \tag{2.35}$$

Gathering (2.34) and (2.35) enables us to conclude.

2.5.4 Proof of Lemma 33

Making the link between ODEs and SDEs (see Doss [33]), one can check that $(\overline{Y}^N_{t_1},\dots,\overline{Y}^N_{t_N})$ has the same distribution law as $(\overline{\overline{Y}}_{2t_1},\dots,\overline{\overline{Y}}_{2t_N})$ where $(\overline{\overline{Y}}_t)_{t\in[0,2T]}$ is solution of the following inhomogeneous SDE $\overline{\overline{Y}}_t = y_0 + \int_0^t \overline{b}(s,\overline{\overline{Y}}_s)ds + \int_0^t \overline{\sigma}(s,\overline{\overline{Y}}_s)dW_s$ with,
$\forall (s,y) \in [0,2T]\times\mathbb{R}$,

$$\overline{b}(s,y) = \begin{cases} b(y) - \frac{1}{2}\sigma\sigma'(y) & \text{if } s \in \bigcup_{k=0}^{N-1}\left[\frac{(4k+1)T}{2N}, \frac{(4k+3)T}{2N}\right] \\ -\frac{1}{2}\sigma\sigma'(y) & \text{otherwise} \end{cases}$$

and

$$\overline{\sigma}(s,y) = \begin{cases} 0 & \text{if } s \in \bigcup_{k=0}^{N-1}\left[\frac{(4k+1)T}{2N}, \frac{(4k+3)T}{2N}\right] \\ \sigma(y) & \text{otherwise} \end{cases}$$

Since these coefficient have a uniform in time linear growth in the spatial variable, one easily concludes.

Chapitre 3

Erreur faible uniforme en temps pour le schéma d'Euler

Dans ce chapitre, on s'intéresse à l'erreur faible trajectorielle du schéma d'Euler. On donne un début de réponse en prouvant que la vitesse de convergence faible est uniforme en temps pour les lois marginales.

3.1 Introduction

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $(W_t)_{t\in[0,T]}$ un mouvement Brownien de dimension $r \geq 1$, muni de sa filtration naturelle $(\mathcal{F}_t)_{t\in[0,T]}$. On considère l'EDS d-dimensionnelle suivante, $d \geq 1$:

$$\begin{cases} dX_t = b(X_t)dt + \sigma(X_t)dW_t \\ X_0 = x \in \mathbb{R}^d \end{cases} \tag{3.1}$$

avec $b : \mathbb{R}^d \to \mathbb{R}^d$ et $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times r}$. On désigne par $(X_t^x)_{t\in[0,T]}$ la solution de (3.1) partant de x et par $(X_t^{x,n})_{t\in[0,T]}$ son schéma d'Euler, n étant le nombre de points de discrétisation de l'intervalle $[0, T]$. L'objectif de cette note est d'estimer l'erreur faible du schéma d'Euler uniformément en temps.

Commençons par introduire les notations que nous allons utiliser par la suite :

- Pour un multi-indice $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$, on note par $|\alpha| = \alpha_1 + \cdots + \alpha_d$ sa longueur et par ∂^α l'opérateur différentiel $\partial^{|\alpha|}/\partial_1^{\alpha_1} \ldots \partial_d^{\alpha_d}$.
- $\mathcal{C}_b^\infty(\mathbb{R}^d)$ désigne l'espace des fonctions infiniment dérivables de $\mathbb{R}^d$ avec des dérivées de tout ordre bornées et $\mathcal{C}_{b\geq 1}^\infty(\mathbb{R}^d)$ désigne l'espace des fonctions $\mathcal{C}^\infty$ qui ont des dérivées de tout ordre ≥ 1 bornées (donc non nécessairement bornées elles-mêmes).
- On désigne par a la matrice $\sigma\sigma^*$ (on note la transposition par une étoile).
- Pour $t \in [0, T]$, on désigne par $\tau_t = \frac{\lfloor n\frac{t}{T}\rfloor}{n}T$ le point de discrétisation qui vient juste avant t.
- Enfin, quand elles existent, on notera par $p(t, x, .)$ et $p_n(t, x, .)$ les densités de X_t^x et de $X_t^{x,n}$ respectivement : $\forall \mathcal{A} \in \mathcal{B}(\mathbb{R}^d), \mathbb{P}\left(X_t^x \in \mathcal{A}\right) = \int_{\mathcal{A}} p(t, x, y)dy$ et $\mathbb{P}\left(X_t^{x,n} \in \mathcal{A}\right) = \int_{\mathcal{A}} p_n(t, x, y)dy$.

L'étude de la convergence du schéma d'Euler a fait l'objet d'une recherche approfondie. Loins d'être exhaustifs, nous citons quelques travaux importants parus dans la littérature :

- Talay et Tubaro [117] ont obtenu un développement en puissances de $\frac{1}{n}$ de l'erreur faible pour des fonctions test $\mathcal{C}^\infty$ à croissance polynômiale et en supposant que les coefficients de l'EDS sont dans $\mathcal{C}^\infty_{b\geq 1}(\mathbb{R}^d)$.
- En utilisant le calcul de Malliavin, Bally et Talay [7] ont généralisé ce résultat à des fonctions test seulement mesurables bornées dans le cas où l'EDS est uniformément hypoelliptique. Dans un deuxième papier (Bally et Taly [8]), ils ont aussi montré que, quand l'EDS est uniformément elliptique, la différence entre la densité de la solution à l'instant terminal et celle du schéma d'Euler admet un développement en puissances de $\frac{1}{n}$ jusqu'à l'ordre 2 avec des majorations gaussiennes des termes de ce développement.
- Sous des hypothèses plus fortes, en l'occurrence que l'EDS est uniformément elliptique et que ses coefficients sont dans $\mathcal{C}^\infty_b(\mathbb{R}^d)$, et en utilisant une approche originale basée sur la méthode paramétrix, Konakov et Mammen [70] ont obtenu un développement à tout ordre de cette différence. Les termes du développement dépendent de n mais sont uniformément contrôlés par des majorations gaussiennes. Une autre méthode alternative à la technique d'analyse d'erreur de Talay et Tubaro [117] a été proposée par Kohatsu-Higa [69] qui analyse l'erreur du schéma d'Euler directement via le calcul de Malliavin, avec des hypothèses de régularité au sens de Malliavin de la solution de l'EDS.
- Les résultats précédents sont valables pour un temps terminal fixé. Kurtz et Protter [75] ont étudié la vitesse de convergence en loi du processus $(X_t^{x,n})_{t\in[0,T]}$ vers $(X_t^x)_{t\in[0,T]}$ et ont montré qu'elle est en $\frac{1}{\sqrt{n}}$. Sous les mêmes hypothèses que Konakov et Mammen [70], Guyon [52] a démontré un développement en puissances de $\frac{1}{n}$ de la différence entre la densité de la solution et celle du schéma à tout instant. Les termes du développement sont contrôlés par des majorations gaussiennes et l'auteur montre aussi comment utiliser son résultat pour contrôler l'erreur faible du schéma d'Euler avec une classe plus large de fonctions test, par exemple avec les distributions tempérées.
- Le développement de la différence entre les densités obtenu par Guyon [52] ne donne pas la bonne asymptotique pour des temps petits. Récemment, Gobet et Labart [50] ont obtenu une majoration plus fine de cette différence dans le cadre plus général d'EDS inhomogènes en temps et sous des hypothèses plus faibles que celles citées précédemment.
  Plus précisément, les auteurs supposent que $b : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ et $\sigma : [0,T]\times\mathbb{R}^{d\times r}$ vérifient les hypothèses suivantes :

  $(\mathcal{H}_{\mathbf{GL}})$

  $$\forall 1\leq i\leq d \text{ et } \forall 1\leq j\leq r, b_i, \sigma_{i,j}\in\mathcal{C}_b^{1,3} \text{ et } \partial_t\sigma_{i,j}\in\mathcal{C}_b^{0,1}$$

  $$\exists\eta>0 \text{ tel que } \forall x,\xi\in\mathbb{R}^d, \xi^* a(x)\xi\geq\eta\|\xi\|^2$$

  où $\mathcal{C}_b^{k,l}$ désigne l'espace des fonctions continûment différentiables qui vont de $[0,T]\times\mathbb{R}^d$ dans $\mathbb{R}$ et qui admettent des dérivées en temps (respectivement en espace) uniformément bornées jusqu'à l'ordre k (respectivement l). Ils obtiennent alors le résultat suivant

Théorème 4 *(Gobet et Labart [50])*
Sous l'hypothèse ($\mathcal{H}_{GL}$), il existe $c>0$ et K une fonction croissante qui dépend uniquement

de la dimension d et des bornes sur les coefficients de l'EDS et de leurs dérivées, tels que

$$\forall (t,x,y) \in]0,T] \times \mathbb{R}^d \times \mathbb{R}^d, \qquad |p(t,x,y) - p_n(t,x,y)| \leq \frac{K(T)T}{n} t^{-\frac{d+1}{2}} \exp\left(-c\frac{\|x-y\|^2}{t}\right)$$

La preuve de ce théorème fait appel au calcul de Malliavin, en particulier à des résultats fins dus à Kusuoka et Stroock [78].

Grâce à ces résultats et à divers autres travaux de recherche, nous avons une connaissance de plus en plus fine de l'erreur faible du schéma d'Euler à un instant donné. En revanche, l'erreur faible trajectorielle reste une question ouverte : pour une fonctionnelle $f : \mathcal{C}([0,T]) \to \mathbb{R}$, quelle est la vitesse de convergence de $\left|\mathbb{E}\left(f\left((X_t^x)_{t\in[0,T]}\right) - f\left((X_t^{x,n})_{t\in[0,T]}\right)\right)\right|$ en fonction du pas de discrétisation ? On peut trouver dans la littérature des travaux qui abordent cette question pour des fonctionnelles particulières, généralement inspirées par des exemples provenant de la finance de marché. Par exemple, Gobet [49] traite le cas des options barrières en montrant que cette vitesse est en $\frac{1}{n}$ pour les fonctionnelles du type $\mathbb{1}_{\{\forall 0\leq t\leq T, X_t^x \in \mathcal{D}\}} f(X_T^x)$ où $\mathcal{D}$ est un domaine ouvert de $\mathbb{R}^d$ et f une fonction dont le support est strictement inclus dans $\mathcal{D}$. L'auteur montre aussi que la version discrète du schéma d'Euler converge en $\frac{1}{\sqrt{n}}$. Temam [121] s'est intéressé aux options asiatiques et a obtenu une vitesse en $\frac{1}{n}$ pour des fonctionnelles du type $f\left(\int_0^T X_t^x dt\right)$ pour f une fonction lipschitzienne. Tanré [120] a montré que c'est également le cas pour des fonctionnelles du type $\int_0^T f(X_t^x)dt$ avec f seulement mesurable bornée. Citons également Seumen Tonou [113] qui s'est intéressé aux options Lookback et qui a obtenu une vitesse en $\frac{1}{\sqrt{n}}$ pour la version discrète du schéma d'Euler.

Ouvrons une petite parenthèse pratique. En finance, le pricing d'options revient souvent au calcul d'une espérance du type $\mathbb{E}\left(f\left((S_t)_{t\in[0,T]}\right)\right)$ où $(S_t)_{t\in[0,T]}$ est la solution d'une équation différentielle stochastique. Quand la fonctionnelle f est seulement fonction de la valeur terminale S_T, on parle d'options vanilla (Calls, Puts, ...). Quand f est une vraie fonctionnelle de la trajectoire, on parle d'options path-dependent (options asiatiques, options lookback, options à barrières, ...). Dans le premier cas, quand on utilise un schéma de discrétisation pour l'EDS vérifiée par $(S_t)_{t\in[0,T]}$, ce qui compte c'est la convergence faible au sens classique. Dans le deuxième cas, le critère le plus pertinent pour comparer différents schémas de discrétisation ce n'est pas la convergence forte comme il est communément admis mais bien la convergence faible trajectorielle.

La convergence forte du schéma d'Euler à la vitesse $\frac{1}{\sqrt{n}}$ nous assure que la vitesse faible trajectorielle est au moins égale à $\frac{1}{\sqrt{n}}$ pour les fonctionnelles lipschitziennes :

$$\begin{aligned}
\left|\mathbb{E}\left(f\left((X_t^x)_{t\in[0,T]}\right) - f\left((X_t^{x,n})_{t\in[0,T]}\right)\right)\right| &\leq \mathbb{E}\left(\left|f\left((X_t^x)_{t\in[0,T]}\right) - f\left((X_t^{x,n})_{t\in[0,T]}\right)\right|\right) \\
&\leq L_f \mathbb{E}\left(\sup_{t\in[0,T]} \|X_t^x - X_t^{x,n}\|\right) \\
&= \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

où L_f désigne la constante de Lipschitz de f.

Cependant, faire passer la valeur absolue à l'intérieur de l'espérance donne une estimation grossière et on peut espérer que, comme pour la vitesse faible classique, la vitesse de convergence est meilleure que $\frac{1}{\sqrt{n}}$.

En réalité, pour des fonctionnelles lipschitziennes, il est plus judicieux de considérer la distance de Wasserstein [1] entre les lois des deux processus $(X_t^x)_{t\in[0,T]}$ et $(X_t^{x,n})_{t\in[0,T]}$. Nous en rappelons la

1. La terminologie varie dans la littérature. On parle aussi de distance de Monge-Kontorovith ou de Kontorovitch-Rubinstein.

définition dans le cadre des espace vectoriels normés (voir par exemple Villani [124] ou Rachev et Rüschendorf [101]) :

Définition 5 — *Soient $(E, \|\cdot\|_E)$ un espace vectoriel normé, μ_X et μ_Y deux lois de probabilité sur E. La distance de Wasserstein entre μ_X et μ_Y est définie par*

$$d_W(\mu_X, \mu_Y) = \inf_{\pi \in \Pi(\mu_X, \mu_Y)} \int_{E^2} \|x - y\|_E \, d\pi(x, y)$$

où $\Pi(\mu_X, \mu_Y)$ désigne l'espace de toutes les mesures de probabilités π sur $E \times E$ qui ont pour marginales μ_X et μ_Y (i.e. $\forall \mathcal{A} \in \mathcal{B}(E)$, $\pi(\mathcal{A} \times E) = \mu_X(\mathcal{A})$ et $\pi(E \times \mathcal{A}) = \mu_Y(\mathcal{A})$). On dit que π réalise un couplage entre μ_X et μ_Y.

Le théorème de duality de Kantorovitch (voir Théorème 2.5.6 page 94 de Rachev et Rüschendorf [101]) donne une formulation alternative de la distance de Wasserstein, plus appropriée à notre contexte :

Proposition 6 — *On peut définir la distance de Wasserstein par*

$$d_W(\mu_X, \mu_Y) = \sup_{\phi \in Lip_1(E)} \left| \int_E \phi(x) d\mu_X(x) - \int_E \phi(y) d\mu_Y(y) \right|$$

où $Lip_1(E) = \left\{\phi : E \to \mathbb{R}; \phi \in L^1(d\mu_X) \cap L^1(d\mu_Y) \text{ et } \forall (x, y) \in E^2, |\phi(x) - \phi(y)| \leq \|x - y\|_E\right\}$.

De plus, le supremum ne change pas si on se restreint aux applications $\phi \in Lip_1(E)$ bornées.

On voit bien que l'étude de la convergence faible trajectorielle du schéma d'Euler revient à préciser le comportement en fonction du pas de discrétisation de $d_W(P_{X^x}, P_{X^{x,n}})$ où P_{X^x} et $P_{X^{x,n}}$ désignent respectivement les lois de $(X_t^x)_{t \in [0,T]}$ et de $(X_t^{x,n})_{t \in [0,T]}$. Pour cela, une première étape consiste à contrôler la distance de Wasserstein entre les marginales de ces processus uniformément en temps. C'est le résultat que l'on se propose de démontrer ci-après.

3.2 Résultat principal

Commençons par spécifier le cadre d'hypothèses sous lequel on va travailler :

($\mathcal{H}$15)

$$\forall 1 \leq i \leq d \text{ et } \forall 1 \leq j \leq r, \quad b_i, \sigma_{i,j} \in \mathcal{C}_b^\infty(\mathbb{R}^d)$$

$$\exists \eta > 0 \text{ tel que } \forall x, \xi \in \mathbb{R}^d, \quad \xi^* a(x) \xi \geq \eta \|\xi\|^2$$

Notre résultat principal est le suivant

Théorème 7 *Sous l'hypothèse ($\mathcal{H}$15), il existe une constante C indépendante de n tel que*

$$\sup_{0 \leq t \leq T} d_W\left(P_{X_t^x}, P_{X_t^{x,n}}\right) \leq \frac{C}{n}$$

où, $\forall t \in [0, T]$, $P_{X_t^x}$ et $P_{X_t^{x,n}}$ désignent respectivement les lois de X_t^x et de $X_t^{x,n}$.

Avant d'en donner la preuve, remarquons que ce théorème peut être vu comme une conséquence directe du résultat de Gobet et Labart [50].

En effet, soit $f : \mathbb{R}^d \to \mathbb{R}$ tel que $f \in Lip_1(\mathbb{R}^d)$. Pour tout $t \in]0, T]$, on a d'après le Théorème 4

$$\begin{aligned} |\mathbb{E}\left(f(X_t^x)\right) - \mathbb{E}\left(f(X_t^{x,n})\right)| &= \left| \int_{\mathbb{R}^d} f(y)(p(t,x,y) - p_n(t,x,y))dy \right| \\ &= \left| \int_{\mathbb{R}^d} (f(y) - f(x))(p(t,x,y) - p_n(t,x,y))dy \right| \\ &\leq \int_{\mathbb{R}^d} \|y - x\| \frac{K(T)T}{n} t^{-\frac{d+1}{2}} \exp\left(-c\frac{\|x-y\|^2}{t}\right) dy \end{aligned}$$

La deuxième égalité vient du fait que $\int_{\mathbb{R}^d}(p(t,x,y) - p_n(t,x,y))dy = 1 - 1 = 0$. En faisant le changement de variables $z = \frac{y-x}{\sqrt{t}}$, on obtient que

$$|\mathbb{E}\left(f(X_t^x)\right) - \mathbb{E}\left(f(X_t^{x,n})\right)| \leq \frac{1}{n} K(T)T \int_{\mathbb{R}^d} \|z\| \exp\left(-c\|z\|^2\right) dz$$

donc il existe une constante $C > 0$ indépendante de t, de n et de f telle que $|\mathbb{E}\left(f(X_t^x)\right) - \mathbb{E}\left(f(X_t^{x,n})\right)| \leq \frac{C}{n}$. On conclut grâce à la Proposition 6.

Cela dit, nous avons obtenu ce résultat indépendamment du travail de Gobet et Labart [50]. À la différence de leur approche, basée sur le calcul de Malliavin, nous avons utilisé une méthode probabiliste/analytique classique.

3.3 Résultats auxiliaires

L'hypothèse ($\mathcal{H}15$) nous assure que les marginales en temps de la solution de l'EDS (3.1) et de son schéma d'Euler possèdent des densités. Commençons par rappeler deux résultats connus sur la régularité et le contrôle des dérivées de ces densités d'une part et sur le contrôle de convolutions en espace particulières qui apparaissent naturellement dans l'étude de l'erreur faible du schéma d'Euler d'autre part.

Les résultats concernant la densité de la solution de l'EDS remontent à Friedman [41] (cf. Théorème 7 page 260). Pour la densité du schéma d'Euler, le résultat est essentiellement dû à Konakov et Mammen [70]. Se référer également au Lemme 16 page 895 de Guyon [52]. Le Lemme 9 ci-dessous est tiré de la Proposition 5 page 884 de Guyon [52].

Lemme 8 — *Sous l'hypothèse ($\mathcal{H}15$), on a*

– $\forall t \in]0, T], p(t, ., .)$ *est* $\mathcal{C}^\infty$ *et* $\forall \alpha, \beta \in \mathbb{N}^d, \exists c_1 \geq 0$ *et* $c_2 > 0$ *tel que* $\forall t \in]0, T]$ *et* $\forall x, y \in \mathbb{R}^d$

$$\left|\partial_x^\alpha \partial_y^\beta p(t,x,y)\right| \leq c_1 t^{-\frac{|\alpha|+|\beta|+d}{2}} \exp\left(-c_2 \frac{\|x-y\|^2}{t}\right) \tag{3.2}$$

$$\left|\partial_x^\alpha p(t,x,x+y\sqrt{t})\right| \leq c_1 t^{-\frac{d}{2}} \exp\left(-c_2\|y\|^2\right) \tag{3.3}$$

– $\forall n \in \mathbb{N}^*, t \in]0, T], p_n(t, ., .)$ *est* $\mathcal{C}^\infty$ *et* $\forall \alpha, \beta \in \mathbb{N}^d, \exists c_1 \geq 0$ *et* $c_2 > 0$ *tel que* $\forall t \in]0, T], \forall x, y \in \mathbb{R}^d$ *et* $\forall n \in \mathbb{N}^*$

$$\left|\partial_x^\alpha \partial_y^\beta p_n(t,x,y)\right| \leq c_1 t^{-\frac{|\alpha|+|\beta|+d}{2}} \exp\left(-c_2 \frac{\|x-y\|^2}{t}\right) \tag{3.4}$$

$$\left|\partial_x^\alpha p_n(t,x,x+y\sqrt{t})\right| \leq c_1 t^{-\frac{d}{2}} \exp\left(-c_2\|y\|^2\right) \tag{3.5}$$

Lemme 9 — *Soit $g \in \mathcal{C}_b^\infty(\mathbb{R}^d)$ et $l \in \mathbb{N}^d$. Sous l'hypothèse 15, la fonction*

$$\begin{array}{rcl} \pi : \{(s,t) \in \mathbb{R}^2; 0 < s < t \leq T\} \times \mathbb{R}^d \times \mathbb{R}^d & \to & \mathbb{R} \\ (s,t,x,y) & \mapsto & \int_{\mathbb{R}^d} g(z) p_n(s,x,z) \partial_x^l p(t-s,z,y) dz \end{array}$$

vérifie

- $\forall 0 < s < t \leq T, \pi(s,t,.,.)$ *est* $\mathcal{C}^\infty$.
- $\forall \alpha, \beta \in \mathbb{N}^d, \exists c_1 \geq 0$ *et* $c_2 > 0$ *tel que* $\forall 0 < s < t \leq T$ *et* $\forall x, y \in \mathbb{R}^d$

$$\left| \partial_x^\alpha \partial_y^\beta \pi(s,t,x,y) \right| \leq c_1 t^{-\frac{|\alpha|+|\beta|+d+|l|}{2}} \exp\left(-c_2 \frac{\|x-y\|^2}{t} \right). \tag{3.6}$$

Dans la preuve du Théorème 7, en plus du précédent lemme, nous aurons besoin d'un autre résultat sur l'estimation de convolutions en espace faisant intervenir la densité du schéma d'Euler :

Proposition 10 — *Soient $h \in \mathcal{C}_b^\infty(\mathbb{R}^d)$ et $g \in \mathcal{C}_{b\geq 1}^\infty(\mathbb{R}^d)$. Sous l'hypothèse ($\mathcal{H}$15), la fonction*

$$\begin{array}{rcl} \pi : \{(s,t) \in \mathbb{R}^2; 0 < \frac{t}{2} < s < t \leq T\} \times \mathbb{R}^d \times \mathbb{R}^d & \to & \mathbb{R} \\ (s,t,x,y) & \mapsto & \int_{\mathbb{R}^d} h(z)(g(z)-g(y)) p_n(s,x,z) p_n(t-s,z,y) dz \end{array}$$

vérifie

i) $\forall 0 < \frac{t}{2} < s < t \leq T, \pi(s,t,.,.)$ *est* $\mathcal{C}^\infty$.

ii) $\forall \alpha, \beta \in \mathbb{N}^d, \exists c_1 \geq 0$ *et* $c_2 > 0$ *tel que* $\forall 0 < \frac{t}{2} < s < t \leq T$ *et* $\forall x, y \in \mathbb{R}^d$

$$\left| \partial_x^\alpha \partial_y^\beta \pi(s,t,x,y) \right| \leq c_1 t^{-\frac{|\alpha|+|\beta|+d-1}{2}} \exp\left(-c_2 \frac{\|x-y\|^2}{t} \right) \tag{3.7}$$

$$\left| \partial_x^\alpha \pi(s,t,x,x+y\sqrt{t}) \right| \leq c_1 t^{-\frac{d-1}{2}} \exp\left(-c_2 \|y\|^2 \right). \tag{3.8}$$

Preuve :

En appliquant le théorème de convergence dominée, on montre facilement la propriété *i)*. Pour démontrer la deuxième propriété, nous reprenons la preuve de Guyon [52] en apportant quelques modifications qui permettent de tirer profit de l'accroissement qui apparaît dans l'intégrale.

Faisons le changement de variables $z = y - \xi\sqrt{t-s}$:

$$\pi(s,t,x,y) = (t-s)^{\frac{d}{2}} \int_{\mathbb{R}^d} h(y-\xi\sqrt{t-s})(g(y-\xi\sqrt{t-s})-g(y)) p_n(s,x,y-\xi\sqrt{t-s}) p_n(t-s,y-\xi\sqrt{t-s},y) d\xi$$

En appliquant la formule de Leibniz, on voit que $\partial_x^\alpha \partial_y^\beta \pi(s,t,x,y)$ est égale à une somme pondérée de termes de la forme $\int_{\mathbb{R}^d} \delta(\xi) d\xi$ où

$$\delta(\xi) = (t-s)^{\frac{d}{2}} \partial_y^{\beta_1} h(y-\xi\sqrt{t-s}) \partial_y^{\beta_2} (g(y-\xi\sqrt{t-s})-g(y)) \partial_x^\alpha \partial_y^{\beta_3} p_n(s,x,y-\xi\sqrt{t-s}) \partial_y^{\beta_4} p_n(t-s,y-\xi\sqrt{t-s},y)$$

avec $|\beta_1| + |\beta_2| + |\beta_3| + |\beta_4| = |\beta|$.

Grâce au Lemme 8, on a

$$|\partial_y^{\beta_4} p_n(t-s, y-\xi\sqrt{t-s}, y)| \leq c_1(t-s)^{-\frac{d}{2}} e^{-c_2\|\xi\|^2}$$

et, du fait que $0 < \frac{t}{2} < s < t \leq T$,

$$\begin{aligned} |\partial_x^\alpha \partial_y^{\beta_3} p_n(s, x, y-\xi\sqrt{t-s})| &\leq c_1 s^{-\frac{|\alpha|+|\beta_3|+d}{2}} e^{-c_2 \frac{\|x-y+\xi\sqrt{t-s}\|^2}{s}} \\ &\leq c_3 t^{-\frac{|\alpha|+|\beta_3|+d}{2}} e^{-c_2 \frac{\|x-y+\xi\sqrt{t-s}\|^2}{t}} \end{aligned}$$

où $c_3 = c_1 2^{\frac{|\alpha|+|\beta_3|+d}{2}}$. Comme $t-s \leq \frac{t}{2}$ on a

$$\begin{aligned} \|\xi\|^2 + \frac{\|x-y+\xi\sqrt{t-s}\|^2}{t} &\geq \frac{\|x-y\|^2}{2t} + \|\xi\|^2(1-\frac{t-s}{t}) \\ &\geq \frac{\frac{\|x-y\|^2}{t} + \|\xi\|^2}{2} \end{aligned}$$

donc, puisque $h \in \mathcal{C}_b^\infty$, il existe $c_4 > 0$ tel que

$$\begin{aligned} &|(t-s)^{\frac{d}{2}} \partial_y^{\beta_1} h(y-\xi\sqrt{t-s})\, \partial_y^{\beta_4} p_n(t-s, y-\xi\sqrt{t-s}, y)\, \partial_x^\alpha \partial_y^{\beta_3} p_n(s, x, y-\xi\sqrt{t-s})| \\ &\qquad \leq c_4 t^{-\frac{|\alpha|+|\beta|+d}{2}} e^{-\frac{c_2}{2}\frac{\|x-y\|^2}{t}} e^{-\frac{c_2}{2}\|\xi\|^2} \end{aligned}$$

Comme $g \in \mathcal{C}_{b\geq 1}^\infty$, il existe $K_{\beta_2} > 0$ tel que $|\partial_y^{\beta_2}(g(y-\xi\sqrt{t-s}) - g(y))| \leq K_{\beta_2}\sqrt{t-s}\|\xi\| \leq K_{\beta_2}\sqrt{t}\|\xi\|$. Donc, il existe $c_5 > 0$ tel que

$$\int_{\mathbb{R}^d} |\delta(\xi)| d\xi \leq c_5 t^{-\frac{|\alpha|+|\beta|+d-1}{2}} e^{-\frac{c_2}{2}\frac{\|x-y\|^2}{t}} \int_{\mathbb{R}^d} \|\xi\| e^{-\frac{c_2}{2}\|\xi\|^2} d\xi$$

Ainsi, on retrouve la première inégalité de la propriété *ii)*. La démonstration de la deuxième inégalité repose sur les mêmes arguments. □

Nous allons aussi avoir besoin du lemme suivant :

Lemme 11 — *Soit $g \in Lip_1(\mathbb{R}^d)$. Sous l'hypothèse ($\mathcal{H}15$), $\exists C > 0$ tel que*

$$\mathbb{E}\left[\left|g(X_t^{x,n}) - g(X_{\tau_t}^{x,n})\right|^2\right] \leq C(t-\tau_t) \quad \forall t \in]0, T]$$

Preuve : Puisque $g \in Lip_1(\mathbb{R}^d)$, on a

$$\begin{aligned} \mathbb{E}\left[\left|g(X_t^{x,n}) - g(X_{\tau_t}^{x,n})\right|^2\right] &\leq \mathbb{E}\left[\left\|X_t^{x,n} - X_{\tau_t}^{x,n}\right\|^2\right] \\ &= \mathbb{E}\left[\left\|\int_{\tau_t}^t b(X_{\tau_t}^{x,n})ds + \sigma(X_{\tau_t}^{x,n})dW_s\right\|^2\right] \\ &\leq 2\left((t-\tau_t)^2 \mathbb{E}\left[\left\|b(X_{\tau_t}^{x,n})\right\|^2\right] + \mathbb{E}\left[\left\|\sigma(X_{\tau_t}^{x,n})(W_t - W_{\tau_t})\right\|^2\right]\right) \end{aligned}$$

On conclut en utilisant l'hypothèse ($\mathcal{H}15$). □

3.4 Preuve du Théorème 7

La preuve du théorème s'articule autour de la proposition ci-dessous dont la preuve est reportée à la section suivante :

Proposition 12 — *Sous l'hypothèse ($\mathcal{H}15$), il existe une constante C indépendante de n telle que*

$$\forall f \in \mathcal{C}^\infty(\mathbb{R}^d) \cap Lip_1(\mathbb{R}^d), \forall t \in [0,T], \quad |\mathbb{E}\left(f(X_t^{x,n}) - f(X_t^x)\right)| \leq \frac{C}{n}.$$

Soit $g \in Lip_1(\mathbb{R}^d)$. On sait qu'on peut approcher cette fonction uniformément par une suite de fonctions $\mathcal{C}^\infty$ ayant la même constante de Lipschitz. En effet, soit $(g_m)_{m\in\mathbb{N}^*}$ la suite de fonctions définie par

$$\forall x \in \mathbb{R}^d, \quad g_m(x) = \int_{\mathbb{R}^d} g(y)\phi_m(x-y)dy$$

où ϕ_m est une fonction positive, $\mathcal{C}^\infty$ à support dans $B(0, \frac{1}{m})$, la boule de $\mathbb{R}^d$ de rayon $\frac{1}{m}$, et qui vérifie $\int_{\mathbb{R}^d} \phi_m(y)dy = 1$. Il est clair alors que, $\forall m \in \mathbb{N}^*, g_m \in \mathcal{C}^\infty(\mathbb{R}^d) \cap Lip_1(\mathbb{R}^d)$. De plus,

$$\forall m > 0, \quad \sup_{x\in\mathbb{R}^d} |g(x) - g_m(x)| \leq \frac{1}{m}.$$

D'après la Proposition 12, il existe une constante C indépendante de n telle que, $\forall t \in [0,T]$,

$$\begin{aligned} |\mathbb{E}\left(g(X_t^x) - g(X_t^{x,n})\right)| &\leq |\mathbb{E}\left(g(X_t^x) - g_m(X_t^x)\right)| + |\mathbb{E}\left(g_m(X_t^x) - g_m(X_t^{x,n})\right)| + |\mathbb{E}\left(g_m(X_t^{x,n}) - g(X_t^{x,n})\right)| \\ &\leq \frac{C}{n} + \frac{2}{m} \end{aligned}$$

On conclut en faisant tendre m vers $+\infty$ et en utilisant la Proposition 6.

3.5 Preuve de la Proposition 12

Soit $f \in \mathcal{C}^\infty(\mathbb{R}^d) \cap Lip_1(\mathbb{R}^d)$. On définit la fonction $u : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ par $u(t,x) = \mathbb{E}\left(f(X_t^x)\right)$.
Il est bien connu que sous l'hypothèse ($\mathcal{H}15$), u est solution de l'EDP suivante

$$\begin{cases} \partial_t u(t,x) = \mathcal{L}u(t,x) \\ u(0,x) = f(x) \end{cases} \tag{3.9}$$

où $\mathcal{L}$ désigne l'opérateur différentiel associé à (3.1) :

$$\mathcal{L} = \frac{1}{2}\sum_{i,j=1}^d a_{i,j}\partial_{i,j}^2 + \sum_{i=1}^d b_i\partial_i.$$

Grâce au lemme suivant, on a un contrôle des dérivées en espace de u :

Lemme 13 — *Sous l'hypothèse ($\mathcal{H}15$), $\exists C > 0$ tel que $\forall \alpha \in \mathbb{N}^*$*

$$|\partial_x^\alpha u(t,x)| \leq C t^{-\frac{\alpha-1}{2}} \quad \forall t \in]0,T], x \in \mathbb{R}^d$$

Preuve : Comme $p(t,x,.)$ est une densité, on a $\partial_x^\alpha \int_{\mathbb{R}^d} p(t,x,y)dy = 0$. On peut donc écrire que

$$\begin{aligned} |\partial_x^\alpha u(t,x)| &= \left| \int_{\mathbb{R}^d} f(y) \partial_x^\alpha p(t,x,y) dy \right| \\ &= \left| \int_{\mathbb{R}^d} (f(y) - f(x)) \partial_x^\alpha p(t,x,y) dy \right| \end{aligned}$$

$f \in \mathcal{C}^\infty(\mathbb{R}^d) \cap Lip_1(\mathbb{R}^d)$ donc $\forall x, y \in \mathbb{R}^d, |f(y) - f(x)| \leq \|y - x\|$. Grâce au Lemme 8 on a

$$\begin{aligned} |\partial_x^\alpha u(t,x)| &\leq \int_{\mathbb{R}^d} \|y - x\| c_1 t^{-\frac{|\alpha|+d}{2}} \exp\left(-c_2 \frac{\|x-y\|^2}{t}\right) dy \\ &= c_1 t^{-\frac{|\alpha|-1}{2}} \int_{\mathbb{R}^d} \|z\| \exp\left(-c_2 \|z\|^2\right) dz \end{aligned}$$

donc $|\partial_x^\alpha u(t,x)| \leq C t^{-\frac{|\alpha|-1}{2}}$ avec $C = c_1\, c_2^{-\frac{d+1}{2}}$. □

Notons que la constante C ne dépend de la fonction f qu'à travers sa constante de Lipschitz, égale à 1 en l'occurrence.

L'erreur faible pour la fonction f à un instant $t \in [0,T]$ donné s'écrit

$$\Delta(t) := \mathbb{E}\left(f(X_t^{x,n}) - f(X_t^x)\right) = \mathbb{E}\left(u(0, X_t^{x,n}) - u(t, X_0^{x,n})\right) = \mathbb{E}\left(\int_0^t du(t-s, X_s^{x,n})\right).$$

En applicant la formule d'Itô et en utilisant le fait que u est solution de l'EDP (3.9), on obtient

$$\begin{aligned} du(t-s, X_s^{x,n}) &= -\partial_t u(t-s, X_s^{x,n}) ds + \sum_{i=1}^d \frac{\partial u}{\partial x_i}(t-s, X_s^{x,n}) \left(b_i(X_{\tau_s}^{x,n}) ds + \sum_{k=1}^r \sigma_{i,k}(X_{\tau_s}^{x,n}) dW_s^k \right) \\ &\quad + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 u}{\partial x_i \partial x_j}(t-s, X_s^{x,n})\, a_{i,j}(X_{\tau_s}^{x,n}) ds \\ &= \Big(\sum_{i=1}^d b_i(X_{\tau_s}^{x,n}) \frac{\partial u}{\partial x_i}(t-s, X_s^{x,n}) + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 u}{\partial x_i \partial x_j}(t-s, X_s^{x,n})\, a_{i,j}(X_{\tau_s}^{x,n}) \\ &\quad - \mathcal{L}u(t-s, X_s^{x,n}) \Big) ds + \sum_{i=1}^d \sum_{k=1}^r \sigma_{i,k}(X_{\tau_s}^{x,n}) \frac{\partial u}{\partial x_i}(t-s, X_s^{x,n}) dW_s^k \end{aligned}$$

Grâce à l'hypothèse ($\mathcal{H}15$) et au Lemme 13, les intégrales stochastiques sont de vraies martingales et on obtient :

$$\Delta(t) = \int_0^t \mathbb{E}\left[\sum_{i=1}^d \left(b_i(X_{\tau_s}^{x,n}) - b_i(X_s^{x,n})\right) \frac{\partial u}{\partial x_i}(t-s, X_s^{x,n}) \right.$$
$$\left. + \frac{1}{2}\sum_{i,j=1}^d \left(a_{i,j}(X_{\tau_s}^{x,n}) - a_{i,j}(X_s^{x,n})\right) \frac{\partial^2 u}{\partial x_i \partial x_j}(t-s, X_s^{x,n})\right] ds$$

soit $|\Delta(t)| \leq \left|\int_0^t \Delta_1(s)ds\right| + \left|\int_0^t \Delta_2(s)ds\right|$ avec

$$\Delta_1(s) = \mathbb{E}\left[\sum_{i=1}^d \left(b_i(X_s^{x,n}) - b_i(X_{\tau_s}^{x,n})\right) \frac{\partial u}{\partial x_i}(t-s, X_s^{x,n})\right]$$

et

$$\Delta_2(s) = \mathbb{E}\left[\frac{1}{2}\sum_{i,j=1}^d \left(a_{i,j}(X_s^{x,n}) - a_{i,j}(X_{\tau_s}^{x,n})\right) \frac{\partial^2 u}{\partial x_i \partial x_j}(t-s, X_s^{x,n})\right].$$

Nous allons contrôler ces deux termes séparément. Dans tout ce qui suit, K représente une constante positive qui peut changer d'une ligne à l'autre mais qui ne dépend ni de $t \in [0,T]$ ni de n.

3.5.1 Estimation de $\left|\int_0^t \Delta_1(s)ds\right|$

Appliquons la formule d'Itô une deuxième fois :

$$\begin{aligned}
\Delta_1(s) = \int_{\tau_s}^s \sum_{i=1}^d \mathbb{E}\Bigg[&(b_i(X_{\tau_s}^{x,n}) - b_i(X_r^{x,n}))\frac{\partial^2 u}{\partial t \partial x_i}(t-r, X_r^{x,n}) \\
&+ \sum_{j=1}^d \left((b_i(X_r^{x,n}) - b_i(X_{\tau_s}^{x,n}))\frac{\partial^2 u}{\partial x_j \partial x_i}(t-r, X_r^{x,n}) + \frac{\partial b_i}{\partial x_j}(X_r^{x,n})\frac{\partial u}{\partial x_i}(t-r, X_r^{x,n})\right) b_j(X_{\tau_s}^{x,n}) \\
&+ \frac{1}{2}\sum_{j,k=1}^d \left((b_i(X_r^{x,n}) - b_i(X_{\tau_s}^{x,n}))\frac{\partial^3 u}{\partial x_k \partial x_j \partial x_i}(t-r, X_r^{x,n}) + \frac{\partial^2 b_i}{\partial x_k \partial x_j}(X_r^{x,n})\frac{\partial u}{\partial x_i}(t-r, X_r^{x,n})\right. \\
&\left. + 2\frac{\partial b_i}{\partial x_j}(X_r^{x,n})\frac{\partial^2 u}{\partial x_k \partial x_i}(t-r, X_r^{x,n})\right) a_{j,k}(X_{\tau_s}^{x,n})\Bigg] dr
\end{aligned}$$

En utilisant l'EDP (3.9), on peut simplifier l'expression de $\Delta_1(s)$ comme suit :

$$\Delta_1(s) = \int_{\tau_s}^s \Delta_1^1(r) + \Delta_1^2(r) + \Delta_1^3(r)dr$$

avec

$$\begin{aligned}
\Delta_1^1(r) &= \sum_{i,j,k=1}^{d} \mathbb{E}\left[\frac{\partial u}{\partial x_i}(t-r, X_r^{x,n})\left(\frac{2b_j(X_{\tau_s}^{x,n}) - b_j(X_r^{x,n})}{d}\frac{\partial b_i}{\partial x_j}(X_r^{x,n}) + \frac{a_{j,k}(X_{\tau_s}^{x,n})}{2}\frac{\partial^2 b_i}{\partial x_j \partial x_k}(X_r^{x,n})\right)\right] \\
\Delta_1^2(r) &= \sum_{i,j,k=1}^{d} \mathbb{E}\left[\frac{\partial^2 u}{\partial x_i \partial x_j}(t-r, X_r^{x,n})\left(\frac{b_k(X_{\tau_s}^{x,n}) - b_k(X_r^{x,n})}{2}\frac{\partial a_{j,i}}{\partial x_k}(X_r^{x,n}) + a_{k,i}(X_{\tau_s}^{x,n})\frac{\partial b_j}{\partial x_k}(X_r^{x,n})\right.\right. \\
&\qquad \left.\left. + \frac{(b_j(X_r^{x,n}) - b_j(X_{\tau_s}^{x,n}))(b_i(X_{\tau_s}^{x,n}) - b_i(X_r^{x,n}))}{d}\right)\right] \\
\Delta_1^3(r) &= \sum_{i,j,k=1}^{d} \mathbb{E}\left[\frac{\partial^3 u}{\partial x_i \partial x_j \partial x_k}(t-r, X_r^{x,n})\left(\frac{(a_{j,k}(X_r^{x,n}) - a_{j,k}(X_{\tau_s}^{x,n}))(b_i(X_{\tau_s}^{x,n}) - b_i(X_r^{x,n}))}{2}\right)\right]
\end{aligned}$$

Nous devons donc contrôler les trois termes suivants

$$\left|\int_0^t \Delta_1(s)ds\right| \leq \left|\int_0^t \int_{\tau_s}^s \Delta_1^1(r)drds\right| + \left|\int_0^t \int_{\tau_s}^s \Delta_1^2(r)drds\right| + \left|\int_0^t \int_{\tau_s}^s \Delta_1^3(r)drds\right|$$

Estimation de $\left|\int_0^t \int_{\tau_s}^s \Delta_1^1(r)drds\right|$

On a, grâce à l'hypothèse ($\mathcal{H}15$) et au Lemme 13,

$$\begin{aligned}
|\Delta_1^1(r)| &\leq \sum_{i,j,k=1}^{d} \mathbb{E}\left(\left|\frac{\partial u}{\partial x_i}(t-r, X_r^{x,n})\right|\left|\frac{2b_j(X_{\tau_s}^{x,n}) - b_j(X_r^{x,n})}{d}\frac{\partial b_i}{\partial x_j}(X_r^{x,n}) + \frac{a_{j,k}(X_{\tau_s}^{x,n})}{2}\frac{\partial^2 b_i}{\partial x_j \partial x_k}(X_r^{x,n})\right|\right) \\
&\leq K
\end{aligned}$$

donc

$$\left|\int_0^t \int_{\tau_s}^s \Delta_1^1(r)drds\right| \leq K\int_0^t (s-\tau_s)ds \leq K\frac{1}{n}$$

Estimation de $\left|\int_0^t \int_{\tau_s}^s \Delta_1^2(r)drds\right|$

$\left|\int_0^t \int_{\tau_s}^s \Delta_1^2(r)drds\right|$ fait intervenir des termes de nature différente :

$$\left|\int_0^t \int_{\tau_s}^s \Delta_1^2(r)drds\right| = \left|\int_0^t \int_{\tau_s}^s \Delta_1^{2,1}(r)drds\right| + \left|\int_0^t \int_{\tau_s}^s \Delta_1^{2,2}(r)drds\right| + \left|\int_0^t \int_{\tau_s}^s \Delta_1^{2,3}(r)drds\right|$$

avec

$$\begin{aligned}
\Delta_1^{2,1}(r) &= \sum_{i,j=1}^{d} \mathbb{E}\left[\frac{\partial^2 u}{\partial x_i \partial x_j}(t-r, X_r^{x,n})\left(b_j(X_r^{x,n}) - b_j(X_{\tau_s}^{x,n})\right)\left(b_i(X_{\tau_s}^{x,n}) - b_i(X_r^{x,n})\right)\right] \\
\Delta_1^{2,2}(r) &= \sum_{i,j,k=1}^{d} \mathbb{E}\left[\frac{\partial^2 u}{\partial x_i \partial x_j}(t-r, X_r^{x,n})\left(\frac{b_k(X_{\tau_s}^{x,n}) - b_k(X_r^{x,n})}{2}\frac{\partial a_{j,i}}{\partial x_k}(X_r^{x,n})\right)\right] \\
\Delta_1^{2,3}(r) &= \sum_{i,j,k=1}^{d} \mathbb{E}\left[\frac{\partial^2 u}{\partial x_i \partial x_j}(t-r, X_r^{x,n})\left(a_{k,i}(X_{\tau_s}^{x,n})\frac{\partial b_j}{\partial x_k}(X_r^{x,n})\right)\right]
\end{aligned}$$

Commençons par le premier terme. On a d'après le Lemme 13

$$\begin{aligned}
\left|\int_0^t \int_{\tau_s}^s \Delta_1^{2,1}(r)dr\right| &\leq \sum_{i,j=1}^d \int_0^t \int_{\tau_s}^s \mathbb{E}\Big[\left|\frac{\partial^2 u}{\partial x_i \partial x_j}(t-r, X_r^{x,n})\right| \\
&\qquad \times \left|(b_j(X_r^{x,n}) - b_j(X_{\tau_s}^{x,n}))(b_i(X_{\tau_s}^{x,n}) - b_i(X_r^{x,n}))\right|\Big] drds \\
&\leq K \sum_{i,j=1}^d \int_0^t \int_{\tau_s}^s \frac{1}{\sqrt{t-r}} \mathbb{E}\left[\left|(b_j(X_r^{x,n}) - b_j(X_{\tau_s}^{x,n}))(b_i(X_{\tau_s}^{x,n}) - b_i(X_r^{x,n}))\right|\right] drds
\end{aligned}$$

En utilisant l'inégalité de Cauchy-Schwartz et le Lemme 11, on obtient

$$\begin{aligned}
\left|\int_0^t \int_{\tau_s}^s \Delta_1^{2,1}(r)dr\right| &\leq K \int_0^t \int_{\tau_s}^s \frac{r-\tau_s}{\sqrt{t-r}} drds \\
&\leq K \int_0^t \frac{(s-\tau_s)^2}{\sqrt{t-s}} drds \\
&\leq K \frac{1}{n^2}
\end{aligned}$$

Pareillement, grâce à l'hypothèse ($\mathcal{H}15$) et aux Lemmes 13 et 11,

$$\begin{aligned}
\left|\int_0^t \int_{\tau_s}^s \Delta_1^{2,2}(r)dr\right| &\leq \sum_{i,j,k=1}^d \int_0^t \int_{\tau_s}^s \mathbb{E}\left[\left|\frac{\partial^2 u}{\partial x_i \partial x_j}(t-r, X_r^{x,n})\right| \left|\frac{b_k(X_{\tau_s}^{x,n}) - b_k(X_r^{x,n})}{2} \frac{\partial a_{j,i}}{\partial x_k}(X_r^{x,n})\right|\right] drds \\
&\leq K \sum_{k=1}^d \int_0^t \int_{\tau_s}^s \frac{1}{\sqrt{t-r}} \mathbb{E}\left[\left|b_k(X_r^{x,n}) - b_k(X_{\tau_s}^{x,n})\right|\right] drds \\
&\leq K \int_0^t \frac{(s-\tau_s)^{\frac{3}{2}}}{\sqrt{t-s}} drds \\
&\leq K \frac{1}{n^{\frac{3}{2}}}
\end{aligned}$$

et

$$\begin{aligned}
\left|\int_0^t \int_{\tau_s}^s \Delta_1^{2,3}(r)dr\right| &\leq \sum_{i,j,k=1}^d \int_0^t \int_{\tau_s}^s \mathbb{E}\left[\left|\frac{\partial^2 u}{\partial x_i \partial x_j}(t-r, X_r^{x,n})\right| \left|a_{k,i}(X_{\tau_s}^{x,n}) \frac{\partial b_j}{\partial x_k}(X_r^{x,n})\right|\right] drds \\
&\leq K \int_0^t \frac{(s-\tau_s)}{\sqrt{t-s}} drds \\
&\leq K \frac{1}{n}
\end{aligned}$$

donc finalement

$$\left|\int_0^t \int_{\tau_s}^s \Delta_1^2(r)drds\right| \leq K\frac{1}{n}.$$

Estimation de $|\int_0^t \int_{\tau_s}^s \Delta_1^3(r)drds|$

Toujours en utilisant l'hypothèse ($\mathcal{H}15$) et les Lemmes 13 et 11, on montre que

$$\left|\int_0^t \int_{\tau_s}^s \Delta_1^3(r)drds\right| \leq K \int_0^t \int_{\tau_s}^s \frac{r-\tau_s}{t-r}drds. \tag{3.10}$$

Par ailleurs,

$$\begin{aligned}
\int_0^t \int_{\tau_s}^s \frac{r-\tau_s}{t-r}dr\,ds &= \sum_{k=0}^{n\tau_t-1} \int_{t_k}^{t_{k+1}} \int_{t_k}^s \frac{r-t_k}{t-r}dr\,ds + \int_{\tau_t}^t \int_{\tau_t}^s \frac{r-\tau_t}{t-r}dr\,ds \\
&\leq \sum_{k=0}^{n\tau_t-1} \int_{t_k}^{t_{k+1}} \int_{t_k}^s \frac{r-t_k}{t_{k+1}-r}dr\,ds + \int_{\tau_t}^t \int_{\tau_t}^s \frac{r-\tau_t}{t-r}dr\,ds \\
&= \sum_{k=0}^{n\tau_t-1} \int_{t_k}^{t_{k+1}} \frac{s-t_k}{t_{k+1}-s} \int_s^{t_{k+1}} dr\,ds + \int_{\tau_t}^t \frac{s-\tau_t}{t-s} \int_s^t dr\,ds \\
&= \sum_{k=0}^{n\tau_t-1} \frac{(t_{k+1}-t_k)^2}{2} + \frac{(t-\tau_t)^2}{2} \\
&\leq K\frac{1}{n}.
\end{aligned}$$

Ainsi, on a démontré qu'il existe une constante positive K indépendante de $t \in [0,T]$ et de n, tel que

$$\left|\int_0^t \Delta_1(s)ds\right| \leq K\frac{1}{n}.$$

3.5.2 Estimation de $\left|\int_0^t \Delta_2(s)ds\right|$

Pour des raisons techniques, nous allons distinguer deux cas suivant que t est plus petit ou plus grand que le pas de discrétisation :

<u>1er cas :</u> $t \leq \frac{T}{n}$

En utilisant les Lemmes 13 et 11, on a

$$\begin{aligned}
\left|\int_0^t \Delta_2(s)ds\right| &\leq \frac{1}{2}\sum_{i,j=1}^d \int_0^t \mathbb{E}\left[\left|\frac{\partial^2 u}{\partial x_i \partial x_j}(t-s, X_s^{x,n})\right| |a_{i,j}(X_s^{x,n}) - a_{i,j}(x)|\right] ds \\
&\leq K\int_0^t \frac{\sqrt{s}}{\sqrt{t-s}}ds \\
&\leq K\frac{1}{n}.
\end{aligned}$$

<u>2ème cas :</u> $t > \frac{T}{n}$

On a

$$\left|\int_0^t \Delta_2(s)ds\right| = \left|\int_0^{\frac{T}{n}} \Delta_2(s)ds\right| + \left|\int_{\frac{T}{n}}^t \Delta_2(s)ds\right|.$$

D'après le cas précédent, il suffit de contrôler le terme $\left|\int_{\frac{T}{n}}^t \Delta_2(s)ds\right|$. On applique la formule d'Itô :

$$
\begin{aligned}
\Delta_2(s) = \frac{1}{2}\sum_{i,j=1}^{d}\int_{\tau_s}^{s}\mathbb{E}\Bigg[&-(a_{i,j}(X_r^{x,n}) - a_{i,j}(X_{\tau_s}^{x,n}))\frac{\partial^3 u}{\partial t\partial x_i\partial x_j}(t-r, X_r^{x,n}) \\
&+\sum_{k=1}^{d}\Bigg((a_{i,j}(X_r^{x,n}) - a_{i,j}(X_{\tau_s}^{x,n}))\frac{\partial^3 u}{\partial x_k\partial x_i\partial x_j}(t-r, X_r^{x,n}) \\
&\quad+\frac{\partial a_{i,j}}{\partial x_k}(X_r^{x,n})\frac{\partial^2 u}{\partial x_i\partial x_j}(t-r, X_r^{x,n})\Bigg) b_k(X_{\tau_s}^{x,n}) \\
&+\frac{1}{2}\sum_{k,l=1}^{d}\Bigg((a_{i,j}(X_r^{x,n}) - a_{i,j}(X_{\tau_s}^{x,n}))\frac{\partial^4 u}{\partial x_k\partial x_l\partial x_i\partial x_j}(t-r, X_r^{x,n}) \\
&\quad+\frac{\partial^2 a_{i,j}}{\partial x_k\partial x_l}(X_r^{x,n})\frac{\partial^2 u}{\partial x_i\partial x_j}(t-r, X_r^{x,n}) + 2\frac{\partial a_{i,j}}{\partial x_k}(X_r^{x,n})\frac{\partial^3 u}{\partial x_l\partial x_i\partial x_j}(t-r, X_r^{x,n})\Bigg) a_{k,l}(X_{\tau_s}^{x,n})\Bigg] dr
\end{aligned}
$$

Après, on utilise l'EDP (3.9) pour se débarrasser de la dérivée en temps :

$$
\begin{aligned}
\frac{\partial^3 u}{\partial t\partial x_i\partial x_j}(t-r, X_r^{x,n}) &= \frac{\partial^2 \mathcal{L}u}{\partial x_i\partial x_j}(t-r, X_r^{x,n}) \\
&= \sum_{k=1}^{d}\Bigg(\frac{\partial^2 b_k}{\partial x_i\partial x_j}(X_r^{x,n})\frac{\partial u}{\partial x_k}(t-r, X_r^{x,n}) + \frac{\partial b_k}{\partial x_j}(X_r^{x,n})\frac{\partial^2 u}{\partial x_i\partial x_k}(t-r, X_r^{x,n}) \\
&\quad+\frac{\partial b_k}{\partial x_i}(X_r^{x,n})\frac{\partial^2 u}{\partial x_j\partial x_k}(t-r, X_r^{x,n}) + b_k(X_r^{x,n})\frac{\partial^3 u}{\partial x_i\partial x_j\partial x_k}(t-r, X_r^{x,n})\Bigg) \\
&\quad+\frac{1}{2}\sum_{k,l=1}^{d}\Bigg(\frac{\partial^2 a_{k,l}}{\partial x_i\partial x_j}(X_r^{x,n})\frac{\partial^2 u}{\partial x_k\partial x_l}(t-r, X_r^{x,n}) + \frac{\partial a_{k,l}}{\partial x_j}(X_r^{x,n})\frac{\partial^3 u}{\partial x_i\partial x_k\partial x_l}(t-r, X_r^{x,n}) \\
&\quad+\frac{\partial a_{k,l}}{\partial x_i}(X_r^{x,n})\frac{\partial^3 u}{\partial x_j\partial x_k\partial x_l}(t-r, X_r^{x,n}) + a_{k,l}(X_r^{x,n})\frac{\partial^4 u}{\partial x_i\partial x_j\partial x_k\partial x_l}(t-r, X_r^{x,n})\Bigg)
\end{aligned}
$$

Après calculs, on obtient

$$
\Delta_2(s) = \int_{\tau_s}^{s}\frac{1}{2}\Delta_2^1(r) + \Delta_2^2(r) + \frac{1}{2}\Delta_2^3(r) + \frac{1}{4}\Delta_2^4(r)dr
$$

avec

$$\Delta_2^1(r) = \sum_{i,j,k=1}^{d} \mathbb{E}\left[\frac{\partial u}{\partial x_k}(t-r, X_r^{x,n})\frac{\partial^2 b_k}{\partial x_i \partial x_j}(X_r^{x,n})\left(a_{i,j}(X_{\tau_s}^{x,n}) - a_{i,j}(X_r^{x,n})\right)\right]$$

$$\Delta_2^2(r) = \sum_{i,j=1}^{d} \mathbb{E}\left[\frac{\partial^2 u}{\partial x_i \partial x_j}(t-r, X_r^{x,n})\left(\sum_{k=1}^{d}\left(\frac{\partial b_j}{\partial x_k}(X_r^{x,n})(a_{i,k}(X_{\tau_s}^{x,n}) - a_{i,k}(X_r^{x,n})) + \frac{b_k(X_{\tau_s}^{x,n})}{2}\frac{\partial a_{i,j}}{\partial x_k}(X_r^{x,n})\right) + \frac{1}{4}\sum_{k,l=1}^{d}\frac{\partial^2 a_{i,j}}{\partial x_k \partial x_l}(X_r^{x,n})\left(2a_{k,l}(X_{\tau_s}^{x,n}) - a_{k,l}(X_r^{x,n})\right)\right)\right]$$

$$\Delta_2^3(r) = \sum_{i,j,k=1}^{d} \mathbb{E}\left[\frac{\partial^3 u}{\partial x_i \partial x_j \partial x_k}(t-r, X_r^{x,n})\left(\left(b_k(X_r^{x,n}) - b_k(X_{\tau_s}^{x,n})\right)\left(a_{i,j}(X_{\tau_s}^{x,n}) - a_{i,j}(X_r^{x,n})\right) + \sum_{l=1}^{d}\left(\frac{\partial a_{j,k}}{\partial x_l}(X_r^{x,n})\left(a_{i,l}(X_{\tau_s}^{x,n}) - a_{i,l}(X_r^{x,n})\right) + \frac{\partial a_{i,j}}{\partial x_l}(X_r^{x,n})a_{k,l}(X_{\tau_s}^{x,n})\right)\right)\right]$$

$$\Delta_2^4(r) = \sum_{i,j,k,l=1}^{d} \mathbb{E}\left[\frac{\partial^4 u}{\partial x_i \partial x_j \partial x_k \partial x_l}(t-r, X_r^{x,n})\left(a_{k,l}(X_r^{x,n}) - a_{k,l}(X_{\tau_s}^{x,n})\right)\left(a_{i,j}(X_{\tau_s}^{x,n}) - a_{i,j}(X_r^{x,n})\right)\right]$$

L'estimation des deux premiers termes se fait comme précédemment. Il reste à contrôler $|\int_{\frac{T}{n}}^t \int_{\tau_s}^s \Delta_2^3(r)drds|$ et $|\int_{\frac{T}{n}}^t \int_{\tau_s}^s \Delta_2^4(r)drds|$.

Estimation de $|\int_{\frac{T}{n}}^t \int_{\tau_s}^s \Delta_2^3(r)drds|$

On a

$$\left|\int_{\frac{T}{n}}^t \int_{\tau_s}^s \Delta_2^3(r)drds\right| \leq \left|\int_{\frac{T}{n}}^t \int_{\tau_s}^s \Delta_2^{3,1}(r)drds\right| + \left|\int_{\frac{T}{n}}^t \int_{\tau_s}^s \Delta_2^{3,2}(r)drds\right| + \left|\int_{\frac{T}{n}}^t \int_{\tau_s}^s \Delta_2^{3,3}(r)drds\right|$$

avec

$$\Delta_2^{3,1}(r) = \sum_{i,j,k=1}^{d} \mathbb{E}\left[\frac{\partial^3 u}{\partial x_i \partial x_j \partial x_k}(t-r, X_r^{x,n})\left(\left(b_k(X_r^{x,n}) - b_k(X_{\tau_s}^{x,n})\right)\left(a_{i,j}(X_{\tau_s}^{x,n}) - a_{i,j}(X_r^{x,n})\right)\right)\right]$$

$$\Delta_2^{3,2}(r) = \sum_{i,j,k,l=1}^{d} \mathbb{E}\left[\frac{\partial^3 u}{\partial x_i \partial x_j \partial x_k}(t-r, X_r^{x,n})\frac{\partial a_{j,k}}{\partial x_l}(X_r^{x,n})a_{i,l}(X_r^{x,n})\right]$$

$$\Delta_2^{3,3}(r) = 2\sum_{i,j,k,l=1}^{d} \mathbb{E}\left[\frac{\partial^3 u}{\partial x_i \partial x_j \partial x_k}(t-r, X_r^{x,n})\frac{\partial a_{i,j}}{\partial x_l}(X_r^{x,n})a_{k,l}(X_{\tau_s}^{x,n})\right]$$

Le premier terme est de même nature que le terme $\Delta_1^3(r)$ traité dans la section 3.5.1.
En notant que $\frac{\partial^3 u}{\partial x_i \partial x_j \partial x_k}(t-r, y) = \int_{\mathbb{R}^d}(f(z) - f(x))\frac{\partial^3 p}{\partial x_i \partial x_j \partial x_k}(t-r, y, z)dz$ (voir preuve du

Lemme 13) et en utilisant le théorème de Fubini, on obtient

$$\begin{aligned}
\left|\int_{\frac{T}{n}}^{t}\int_{\tau_s}^{s}\Delta_2^{3,2}(r)drds\right| &\leq \sum_{i,j,k,l=1}^{d}\left|\int_{\frac{T}{n}}^{t}\int_{\tau_s}^{s}\mathbb{E}\left[\frac{\partial^3 u}{\partial x_i\partial x_j\partial x_k}(t-r,X_r^{x,n})\frac{\partial a_{j,k}}{\partial x_l}(X_r^{x,n})a_{i,l}(X_r^{x,n})\right]drds\right| \\
&= \sum_{i,j,k,l=1}^{d}\left|\int_{\frac{T}{n}}^{t}\int_{\tau_s}^{s}\left(\int_{\mathbb{R}^d}\frac{\partial^3 u}{\partial x_i\partial x_j\partial x_k}(t-r,y)\frac{\partial a_{j,k}}{\partial x_l}(y)a_{i,l}(y)p_n(r,x,y)dy\right)drds\right| \\
&= \sum_{i,j,k,l=1}^{d}\left|\int_{\frac{T}{n}}^{t}\int_{\tau_s}^{s}\left(\int_{\mathbb{R}^d}(f(z)-f(x))\pi(r,t,x,z)dz\right)drds\right|
\end{aligned}$$

où $\pi(r,t,x,z)=\displaystyle\int_{\mathbb{R}^d}\frac{\partial a_{j,k}}{\partial x_l}(y)a_{i,l}(y)p_n(r,x,y)\frac{\partial^3 p}{\partial x_i\partial x_j\partial x_k}(t-r,y,z)dy.$
Grâce au Lemme 9, il vient que

$$\begin{aligned}
\left|\int_{\frac{T}{n}}^{t}\int_{\tau_s}^{s}\Delta_2^{3,2}(r)drds\right| &\leq \sum_{i,j,k,l=1}^{d}\int_{\frac{T}{n}}^{t}\int_{\tau_s}^{s}\int_{\mathbb{R}^d}|f(z)-f(x)|\,|\pi(r,t,x,z)|\,dzdrds \\
&\leq K\int_{\frac{T}{n}}^{t}\int_{\tau_s}^{s}\int_{\mathbb{R}^d}\|z-x\|c_1t^{-\frac{d+3}{2}}e^{-c_2\frac{\|z-x\|^2}{t}}dzdrds \\
&\leq K\int_{\frac{T}{n}}^{t}(s-\tau_s)\frac{1}{t}\int_{\mathbb{R}_d}\|w\|e^{-\|w\|^2}dwds \\
&\leq K\frac{1}{n}.
\end{aligned}$$

Regardons maintenant le dernier terme. On a

$$\begin{aligned}
\left|\int_{\frac{T}{n}}^{t}\int_{\tau_s}^{s}\Delta_2^{3,3}(r)drds\right| \leq 2&\left|\int_{\frac{T}{n}}^{t}\int_{\tau_s}^{s}\sum_{i,j,k,l=1}^{d}\mathbb{E}\left[\frac{\partial^3 u}{\partial x_i\partial x_j\partial x_k}(t-r,X_r^{x,n})\frac{\partial a_{i,j}}{\partial x_l}(X_r^{x,n})a_{k,l}(X_r^{x,n})\right]drds\right| \\
+2&\left|\int_{\frac{T}{n}}^{t}\int_{\tau_s}^{s}\sum_{i,j,k,l=1}^{d}\mathbb{E}\left[\frac{\partial^3 u}{\partial x_i\partial x_j\partial x_k}(t-r,X_r^{x,n})\frac{\partial a_{i,j}}{\partial x_l}(X_r^{x,n})\left(a_{k,l}(X_{\tau_s}^{x,n})-a_{k,l}(X_r^{x,n})\right)\right]drds\right|
\end{aligned}$$

Le deuxième terme de la somme est de même nature que $\Delta_2^{3,2}(r)$. Il suffit donc de contrôler le terme $\epsilon := \left|\int_{\frac{T}{n}}^{t}\int_{\tau_s}^{s}\sum_{i,j,k,l=1}^{d}\mathbb{E}\left[\frac{\partial^3 u}{\partial x_i\partial x_j\partial x_k}(t-r,X_r^{x,n})\frac{\partial a_{i,j}}{\partial x_l}(X_r^{x,n})\left(a_{k,l}(X_{\tau_s}^{x,n})-a_{k,l}(X_r^{x,n})\right)\right]drds\right|$.
On a

$$\epsilon=\left|\sum_{i,j,k,l=1}^{d}\int_{\frac{T}{n}}^{t}\int_{\tau_s}^{s}\left(\int_{\mathbb{R}^d}\frac{\partial^3 u}{\partial x_i\partial x_j\partial x_k}(t-r,y)\frac{\partial a_{i,j}}{\partial x_l}(y)\pi(\tau_s,r,x,y)dy\right)drds\right|$$

où $\pi(\tau_s,r,x,y)=\displaystyle\int\left(a_{k,l}(z)-a_{k,l}(y)\right)p_n(\tau_s,x,z)p_n(r-\tau_s,z,y)dz$. Donc

$$\epsilon = \left|\sum_{i,j,k,l=1}^{d}\int_{\frac{T}{n}}^{t}\int_{\tau_s}^{s}\int_{\mathbb{R}^d}(f(w)-f(x))\underbrace{\left(\int_{\mathbb{R}^d}\frac{\partial a_{i,j}}{\partial x_l}(y)\pi(\tau_s,r,x,y)\frac{\partial^3 p}{\partial x_i\partial x_j\partial x_k}(t-r,y,w)dy\right)}_{\delta_{\tau_s}(r,t,x,w)}dwdrds\right|$$

Grâce à la Proposition 10, on peut adapter le résultat de la Proposition 5 p. 884 de Guyon [52] pour obtenir l'existence de deux constantes $c_1 \geq 0$ et $c_2 > 0$ indépendantes de τ_s tel que $\forall 0 < \frac{r}{2} < \tau_s < r < t \leq T$ et $\forall x, w \in \mathbb{R}$

$$|\delta_{\tau_s}(r,t,x,w)| \leq c_1 t^{-\frac{d+2}{2}} e^{-c_2 \frac{\|x-w\|^2}{t}}$$

En effet, il suffit de faire le changement de variables $y = x + \sqrt{r}z$ si $r \leq \frac{t}{2}$ et $y = w - \sqrt{t-r}z$ sinon et ensuite procéder comme dans la preuve de la Proposition 10.
Ainsi,

$$\begin{aligned}
\epsilon &\leq \sum_{i,j,k,l=1}^{d} \int_{\frac{T}{n}}^{t} \int_{\tau_s}^{s} \int_{\mathbb{R}^d} |f(w) - f(x)| \, |\delta_{\tau_s}(r,t,x,w)| \, dwdrds \\
&\leq K \int_{\frac{T}{n}}^{t} \int_{\tau_s}^{s} \int_{\mathbb{R}^d} \frac{\|w-x\|}{t^{\frac{d+2}{2}}} e^{-c_2 \frac{\|x-w\|^2}{t}} dwdrds \\
&\leq K \int_{\frac{T}{n}}^{t} (s-\tau_s) \frac{1}{\sqrt{t}} \int_{\mathbb{R}^d} \|v\| e^{-c_2 \|v\|^2} dvds \\
&\leq K \frac{1}{n}.
\end{aligned}$$

Estimation de $|\int_{\frac{T}{n}}^{t} \int_{\tau_s}^{s} \Delta_2^4(r) drds|$

On a

$$\begin{aligned}
\left| \int_{\frac{T}{n}}^{t} \int_{\tau_s}^{s} \Delta_2^4(r) drds \right| &\leq \left| \int_{\frac{T}{n}}^{t} \int_{\tau_s}^{s} \sum_{i,j,k,l=1}^{d} \mathbb{E}\left[\frac{\partial^4 u}{\partial x_i \partial x_j \partial x_k \partial x_l}(t-r, X_r^{x,n}) a_{k,l}(X_{\tau_s}^{x,n}) \left(a_{i,j}(X_{\tau_s}^{x,n}) - a_{i,j}(X_r^{x,n}) \right) \right] \right| \\
&\quad + \left| \int_{\frac{T}{n}}^{t} \int_{\tau_s}^{s} \sum_{i,j,k,l=1}^{d} \mathbb{E}\left[\frac{\partial^4 u}{\partial x_i \partial x_j \partial x_k \partial x_l}(t-r, X_r^{x,n}) a_{k,l}(X_r^{x,n}) \left(a_{i,j}(X_{\tau_s}^{x,n}) - a_{i,j}(X_r^{x,n}) \right) \right] \right| \\
&\leq \sum_{i,j,k,l=1}^{d} \left(\left| \int_{\frac{T}{n}}^{t} \int_{\tau_s}^{s} \int_{\mathbb{R}^d} \frac{\partial^4 u}{\partial x_i \partial x_j \partial x_k \partial x_l}(t-r,y) \pi_{\tau_s}^1(r,x,y) dydrds \right| \right. \\
&\qquad \left. + \left| \int_{\frac{T}{n}}^{t} \int_{\tau_s}^{s} \int_{\mathbb{R}^d} \frac{\partial^4 u}{\partial x_i \partial x_j \partial x_k \partial x_l}(t-r,y) a_{k,l}(y) \pi_{\tau_s}^2(r,x,y) dydrds \right| \right) \\
&= \sum_{i,j,k,l=1}^{d} \left(\left| \int_{\frac{T}{n}}^{t} \int_{\tau_s}^{s} \int_{\mathbb{R}^d} (f(w) - f(x)) \left(\int_{\mathbb{R}^d} \pi^1(\tau_s, r, x, y) \frac{\partial^4 p}{\partial x_i \partial x_j \partial x_k \partial x_l}(t-r,y,w) dy \right) dwdrds \right| \right. \\
&\quad \left. + \left| \int_{\frac{T}{n}}^{t} \int_{\tau_s}^{s} \int_{\mathbb{R}^d} (f(w) - f(x)) \left(\int_{\mathbb{R}^d} a_{k,l}(y) \pi^2(\tau_s, r, x, y) \frac{\partial^4 p}{\partial x_i \partial x_j \partial x_k \partial x_l}(t-r,y,w) dy \right) dwdrds \right| \right)
\end{aligned}$$

avec

$$\begin{aligned}
\pi^1(\tau_s, r, x, y) &= \int_{\mathbb{R}^d} a_{k,l}(z)(a_{k,l}(z) - a_{k,l}(y)) p_n(\tau_s, x, z) p_n(r - \tau_s, z, y) dz \\
\pi^2(\tau_s, r, x, y) &= \int_{\mathbb{R}^d} (a_{k,l}(z) - a_{k,l}(y)) p_n(\tau_s, x, z) p_n(r - \tau_s, z, y) dz.
\end{aligned}$$

De même que précédemment, en utilisant la Proposition 10 pour contrôler $\pi^1_{\tau_s}(r,x,y)$ et $\pi^2_{\tau_s}(r,x,y)$ et en adaptant la proposition 5 p. 884 de Guyon [52], on montre qu'il existe deux constantes $c_1 \geq 0$ et $c_2 > 0$ tel que $\forall 0 < \frac{r}{2} < \tau_s < r < t \leq T$ et $\forall x, w \in \mathbb{R}$

$$\left| \int_{\mathbb{R}^d} \pi^1(\tau_s, r, x, y) \frac{\partial^4 p}{\partial x_i \partial x_j \partial x_k \partial x_l}(t-r, y, w) dy \right| \leq c_1 t^{-\frac{d+3}{2}} e^{-c_2 \frac{\|x-w\|^2}{t}}$$
$$\left| \int_{\mathbb{R}^d} a_{k,l}(y) \pi^2(\tau_s, r, x, y) \frac{\partial^4 p}{\partial x_i \partial x_j \partial x_k \partial x_l}(t-r, y, w) dy \right| \leq c_1 t^{-\frac{d+3}{2}} e^{-c_2 \frac{\|x-w\|^2}{t}}.$$

D'où,

$$\begin{aligned}
\left| \int_{\frac{T}{n}}^t \int_{\tau_s}^s \Delta_2^4(r) dr ds \right| &\leq K \int_{\frac{T}{n}}^t \int_{\tau_s}^s \int_{\mathbb{R}^d} \|w-x\| \frac{c_1}{t^{\frac{d+3}{2}}} e^{-c_2 \frac{\|x-w\|^2}{t}} dw dr ds| \\
&\leq K \int_{\frac{T}{n}}^t (s-\tau_s) \frac{1}{t} \int_{\mathbb{R}^d} \|v\| e^{-c_2 \|v\|^2} dv ds \\
&\leq K \frac{1}{n}
\end{aligned}$$

Ainsi, on a montré qu'il existe une constance C indépendante de n telle que $\forall t \in [0,T], |\Delta(t)| \leq \frac{C}{n}$. En remarquant que cette constante ne dépend de la fonction f qu'à travers sa constante de Lipschitz, on en déduit la Proposition 12.

Deuxième partie

Modélisation de la dépendance en finance : modèle d'indices boursiers et modèles de portefeuilles de crédit

Chapitre 4

Un modèle couplant indice et actions

Ce chapitre reprend un article écrit avec mon directeur de thèse Benjamin Jourdain, soumis pour publication.

Abstract. In this paper, we are interested in continuous time models in which the index level induces some feedback on the dynamics of its composing stocks. More precisely, we propose a model in which the log-returns of each stock may be decomposed into a systemic part proportional to the log-returns of the index plus an idiosyncratic part. We show that, when the number of stocks in the index is large, this model may be approximated by a local volatility model for the index and a stochastic volatility model for each stock with volatility driven by the index. We address calibration of both the limit and the original models.

Introduction

From the early eighties, when trading on stock index was introduced, quantitative finance faced the problem of efficiently pricing and hedging index options along with their underlying components. Many advances have been made for single stock modeling and a variety of solutions to escape from the very restrictive Black & Scholes model has been deeply investigated (such as local volatility models, models with jumps or stochastic volatility models). However, when the number of underlyings is large, index option pricing, or more generally basket option pricing, remains a challenge unless one simply assumes constantly correlated dynamics for the stocks. The problem then is the impossibility of fitting both the stocks and the index smiles.

We try to address this issue by making the dynamics of the stocks depend on the index. The natural fact that the volatility of the index is related to the volatilities of its underlying components

has already been accounted for in the works of Avellaneda *et al.* [5] and Lee *et al.* [82]. In the first paper, the authors use a large deviation asymptotics to reconstruct the local volatility of the index from the local volatilities of the stocks. They express this dependence in terms of implied volatilities using the results of Berestycki *et al.* [10, 11]. In the second paper, the authors reconstruct the Gram-Charlier expansion of the probability density of the index from the stocks using a moments-matching technique. Both papers consider local volatility models for the stocks and a constant correlation matrix but the generalization to stochastic volatility models or to varying correlation coefficients is not straightforward.

Another point of view is to say that the volatility of a composing stock should be related to the index level, or say to the volatility of the index, in some way. This is not astonishing since the index represents the move of the market and reflects the view of the investors on the state of the economy. Moreover, it is coherent with equilibrium economic models like CAPM. Following this idea, we propose a new modeling framework in which the volatility of the index and the volatilities of the stocks are related. We show that, when the number of underlying stocks tends to infinity, our model reduces to a local volatility model for the index and to a stochastic volatility model for the stocks where the stochastic volatility depends on the index level. This asymptotics is reasonable since the number of underlying stocks is usually large. As a consequence, the correlation matrix between the stocks in our model is not constant but stochastic and we show that it is coherent with empirical studies. Finally, we address calibration issues and we show that it is possible, within our framework, to fit both index and stocks smiles. The method we introduce is based on the simulation of SDEs nonlinear in the sense of McKean, and non-parametric estimation of conditional expectations.

This paper is organized as follows. In Section 1, we specify our model for the index and its composing stocks and in Section 2 we study the limiting model when the number of underlying stocks goes to infinity. Section 3 is devoted to calibration issues. Numerical results are presented in Section 4 and the conclusion is given in Section 5.

Acknowledgements: We thank Lorenzo Bergomi, Julien Guyon and all the equity quantitative research team of Societe Generale CIB for numerous fruitful discussions and for providing us with the market data.

4.1 Model Specification

An index is a collection of stocks that reflects the performance of a whole stock market or a specific sector of a market. It is valued as a weighted sum of the value of its underlying components. More precisely, if I_t^M stands for the value at time t of an index composed of M underlyings, then

$$I_t^M = \sum_{j=1}^{M} w_j S_t^{j,M}, \tag{4.1}$$

where $S_t^{j,M}$ is the value of the stock j at time t and the weightings $(w_j)_{j=1\dots M}$ are given constants [1].

Unless otherwise stated, we always work under the risk-neutral probability measure. In order to account for the influence of the index on its underlying components, we specify the following stochastic differential equations for the stocks

$$\forall j \in \{1,\dots,M\}, \quad \frac{dS_t^{j,M}}{S_t^{j,M}} = (r-\delta_j)dt + \beta_j\,\sigma(t,I_t^M)dB_t + \eta_j(t,S_t^{j,M})dW_t^j \tag{4.2}$$

where

- r is the short interest rate.
- $\delta_j \in [0,\infty[$ is the continuous dividend rate of the stock j.
- β_j is the usual beta coefficient of the stock j that quantifies the sensitivity of the stock returns to the index returns (see the seminal paper of Sharpe [114]). It is defined as $\frac{Cov(r_j,r_I)}{Var(r_I)}$ where r_j (respectively r_I) is the rate of return of the stock j (respectively of the index).
- $(B_t)_{t\in[0,T]}, (W_t^1)_{t\in[0,T]}, \dots, (W_t^M)_{t\in[0,T]}$ are independent Brownian motions.
- The coefficients $\sigma, \eta_1, \dots, \eta_M$ satisfy the usual Lipschitz and growth assumptions that ensure existence and strong uniqueness of the solutions (see for example Theorem 5.2.9 of Karatzas and Shreve [63]) :

  ($\mathcal{H}$16) $\exists K$ such that $\forall (t,s_1,s_2) \in [0,T]\times\mathbb{R}^M\times\mathbb{R}^M$,

$$\sum_{j=1}^M \left| s_1^j \sigma\left(t, \sum_{k=1}^M w_k s_1^k\right)\right| + \left|s_1^j \eta_j(t,s_1^j)\right| \le K\left(1+|s_1|\right)$$
$$\sum_{j=1}^M \left| s_1^j \sigma\left(t, \sum_{k=1}^M w_k s_1^k\right) - s_2^j \sigma\left(t, \sum_{k=1}^M w_k s_2^k\right)\right| \le K|s_1-s_2|$$
$$\sum_{j=1}^M \left|s_1^j \eta_j(t,s_1^j) - s_2^j \eta_j(t,s_2^j)\right| \le K|s_1-s_2|$$

As a consequence, the index satisfies the following stochastic differential equation :

$$dI_t^M = rI_t^M dt - \left(\sum_{j=1}^M \delta_j w_j S_t^{j,M}\right)dt + \left(\sum_{j=1}^M \beta_j w_j S_t^{j,M}\right)\sigma(t,I_t^M)dB_t + \sum_{j=1}^M w_j S_t^{j,M}\eta_j(t,S_t^{j,M})dW_t^j \tag{4.3}$$

Before going any further, let us make some preliminary remarks on this framework.

- We have M coupled stochastic differential equations. The dynamics of a given stock depends on all the other stocks composing the index through the volatility term $\sigma(t,I_t^M)$.
- Accounting for the dividends is not relevant for all types of indices. Indeed, for many performance-based indices (such as the German DAX index) dividends and other events are rolled into the final value of the index.
- The cross-correlations between stocks are not constant but stochastic :

$$\rho_{ij} = \frac{\beta_i\beta_j\sigma^2(t,I_t^M)}{\sqrt{\beta_i^2\sigma^2(t,I_t^M)+\eta_i^2(t,S_t^{i,M})}\sqrt{\beta_j^2\sigma^2(t,I_t^M)+\eta_j^2(t,S_t^{j,M})}}$$

1. In most cases, the weightings are either proportional to stock prices or to market capitalization (stock price $\times$ number of shares outstanding) and they are periodically updated but, as usually assumed, we suppose that, up to maturities of the options considered, they do not evolve in time.

Note that they depend not only on the stocks but also on the index. More importantly, it is commonly observed that the more the market is volatile, the more the stocks tend to be highly correlated. This feature is recovered by our model: one can easily check that an increase in the index volatility, with everything else left unchanged, produces an increase in the cross-correlations.
In a recent paper, Cizeau *et al.* [22] show that it is possible to capture the essential features of stocks cross-correlations, in particular in extreme market conditions, by a simple non-Gaussian one factor model. The authors successfully compare different empirical measures of correlation with the prediction of the following model :

$$r_j(t) = \beta_j r_I(t) + \epsilon_j(t) \tag{4.4}$$

where $r_j(t) = \frac{S_t^j}{S_{t-1}^j} - 1$ is the daily return of stock j, $r_I(t)$ is the daily return of the market and the residuals $\epsilon_j(t)$ are independent random variables following a fat-tailed distribution [2]. Our model is in line with (4.4). Indeed, since the beta coefficients are usually narrowly distributed around 1, the factor $\sum_{j=1}^M \beta_j w_j S_t^{j,M}$ of $\sigma(t, I_t^M)$ in (4.3) is close to I_t^M. Moreover, in the next section we show that, for a large number of underlying stocks, one can neglect the term $\sum_{j=1}^M w_j S_t^j \eta_j(t, S_t^j) dW_t^j$ in the dynamics of the index. Hence, if we denote by r_j the log-return of the stock j and by r_{I^M} the log-return of the index, both on a daily basis, we will have

$$r_j = \beta_j r_{I^M} + \eta_j \Delta W^j + \text{drift},$$

where ΔW^j is an independent Gaussian noise. Consequently, in our model too, the return of a stock is decomposed into a systemic part driven by the index, which represents the market, and a residual part.

4.2 Asymptotics for a large number of underlying stocks

The number of underlying components of an index is usually large [3]. It is then meaningful to let M tend to infinity. Since the Brownian motions $(W^j)_{j=1\dots M}$ are independent, one can expect that their contribution to the dynamics governing the index is not significant and drop the corresponding terms in the stochastic differential equation (4.3) which will drastically simplify the model. The aim of this section is to quantify the error we commit by doing so.

To be specific, consider the limit candidate $(I_t)_{t\in[0,T]}$ solution of the following SDE :

$$\begin{cases} dI_t &= (r-\delta) I_t dt + \beta I_t \sigma(t, I_t) dB_t \\ I_0 &= I_0^M \end{cases} \tag{4.5}$$

with δ and β two constant parameters that will be discussed later.

In the following theorem, we give an upper bound for the L^{2p}-distance between $(I_t^M)_{t\in[0,T]}$ and $(I_t)_{t\in[0,T]}$ under mild assumption on the volatility coefficients :

2. The authors have chosen a Student distribution in their numerical experiments.
3. 500 stocks for the S&P 500 index, 100 stocks for the FTSE 100 index, 40 stocks for the CAC40 index, etc.

Theorem 35 — *Let $p \in \mathbb{N}^*$. Under assumption ($\mathcal{H}16$) and if the following assumptions on the volatility coefficients hold,*

($\mathcal{H}17$) $\exists K_b$ *such that* $\forall (t,s) \in [0,T] \times \mathbb{R}_+, \quad |\sigma(t,s)| + |\eta_j(t,s)| \leq K_b$.

($\mathcal{H}18$) $\exists K_\sigma$ *such that* $\forall (t,s_1,s_2) \in [0,T] \times \mathbb{R}_+ \times \mathbb{R}_+, \quad |s_1\sigma(t,s_1) - s_2\sigma(t,s_2)| \leq K_\sigma |s_1 - s_2|$.

then

$$\mathbb{E}\left(\sup_{0\leq t\leq T} |I_t^M - I_t|^{2p}\right) \leq C_T \left(\left(\sum_{j=1}^M w_j^2\right)^p + \left(\sum_{j=1}^M w_j|\beta_j - \beta|\right)^{2p} + \left(\sum_{j=1}^M w_j|\delta_j - \delta|\right)^{2p}\right)$$

where

$$C_T = 8^{2p-1}T^p(T^p + K_pK_b^{2p})C_p \exp\left(4^{2p-1}T(2^{2p-1}K_pT^{p-1}(\beta K_\sigma)^{2p} + (2T)^{2p-1}\delta^{2p} + r^{2p}\,T^{2p-1})\right)$$

and

$$C_p = \max_{1\leq j\leq M} |S_0^{j,M}|^{2p} \exp\left(\left(2r + (2p-1)(\max_{j\geq 1}\beta_j^2 + 1)K_b^2\right)pT\right).$$

The next theorem states that, under an additional assumption on the volatility coefficients, the L^{2p}-distance between a stock $(S_t^{j,M})_{t\in[0,T]}$ and the solution of the SDE obtained by replacing I^M by I

$$\frac{dS_t^j}{S_t^j} = (r - \delta_j)dt + \beta_j\,\sigma(t,I_t)dB_t + \eta_j(t,S_t^j)dW_t^j,\ S_0^j = S_0^{j,M}$$

is controlled by the L^{2p}-distance between I^M and I :

Theorem 36 — *Let $p \in \mathbb{N}^*$. Under the assumptions of Theorem 35 and if*

($\mathcal{H}19$) $\exists K_\eta$ *such that* $\forall (t,s_1,s_2) \in [0,T] \times \mathbb{R}_+ \times \mathbb{R}_+, \quad |s_1\eta(t,s_1) - s_2\eta(t,s_2)| \leq K_\eta|s_1 - s_2|$.
$\exists K_{Lip}$ *such that* $\forall (t,s_1,s_2) \in [0,T] \times \mathbb{R}_+ \times \mathbb{R}_+, \quad |\sigma(t,s_1) - \sigma(t,s_2)| \leq K_{Lip}|s_1 - s_2|$.

Then, $\forall j \in \{1,\ldots,M\}$,

$$\mathbb{E}\left(\sup_{0\leq t\leq T} |S_t^{j,M} - S_t^j|^{2p}\right) \leq \widetilde{C}_T^j \left(\left(\sum_{j=1}^M w_j^2\right)^p + \left(\sum_{j=1}^M w_j|\beta_j - \beta|\right)^{2p} + \left(\sum_{j=1}^M w_j|\delta_j - \delta|\right)^{2p}\right)$$

where

$$\widetilde{C}_T^j = 6^{2p-1}K_pT^p\beta_j^{2p}C_{2p}^{\frac{1}{2}}K_{Lip}^{2p}\ e^{3^{2p-1}((r-\delta_j)^{2p}T^{2p-1} + K_pT^{p-1}K_\eta^{2p} + 2^{2p-1}K_pT^{p-1}\beta_j^{2p}K_b^{2p})T}.$$

Moreover, for $\overline{I}_t^M = \sum_{j=1}^M w_jS_t^j$, *one has*

$$\mathbb{E}\left(\sup_{0\leq t\leq T} |I_t^M - \overline{I}_t^M|^{2p}\right) \leq \widetilde{C}_T \left(\sum_{j=1}^M w_j\right)^{2p} \left(\left(\sum_{j=1}^M w_j^2\right)^p + \left(\sum_{j=1}^M w_j|\beta_j - \beta|\right)^{2p} + \left(\sum_{j=1}^M w_j|\delta_j - \delta|\right)^{2p}\right)$$

where $\widetilde{C}_T = \max_{1\leq j\leq M} \widetilde{C}_T^j$.

The proof for these two theorems can be found in the appendix. Note that, Theorems 35 and 36 yield that I^M is also close to I. In the following corollary, we make explicit the dependence of the coefficients on M and we consider the limit $M \to \infty$:

Corollary 37 — *Under the assumptions of Theorems 35 and 36 and if*

(H20) *there exists a constant A independent of M such that* $\max_{j\geq 1}\left((S_0^{j,M})^2 + (\beta_j^M)^2 + (\delta_j^M)^2\right) \leq A$,

(H21) $P_w^M = \sqrt{\sum_{j=1}^{M}(w_j^M)^2} \underset{M\to\infty}{\longrightarrow} 0$,

(H22) $P_\beta^M = \sum_{j=1}^{M} w_j^M |\beta_j^M - \beta| \underset{M\to\infty}{\longrightarrow} 0$,

(H23) $P_\delta^M = \sum_{j=1}^{M} w_j^M |\delta_j^M - \delta| \underset{M\to\infty}{\longrightarrow} 0$,

then one has

$$\mathbb{E}\left(\sup_{0\leq t\leq T}|I_t^M - I_t|^2\right) \underset{M\to\infty}{\longrightarrow} 0$$

and

$$\forall j \in \{1,\dots,M\}, \quad \mathbb{E}\left(\sup_{0\leq t\leq T}|S_t^{j,M} - S_t^j|^2\right) \underset{M\to\infty}{\longrightarrow} 0.$$

If, in addition, $\sup_M \sum_{j=1}^{M} w_j^M < \infty$ *then*

$$\mathbb{E}\left(\sup_{0\leq t\leq T}|I_t^M - \overline{I}_t^M|^2\right) \underset{M\to\infty}{\longrightarrow} 0.$$

Let us briefly comment on these additional assumptions :

- Assumption ($\mathcal{H}20$) is a technical assumption that prevents the constants C_T and $\widetilde{C}_T$ appearing in the Theorems 35 and 36 from depending on M. It says that the initial stock levels, the beta coefficients and the dividend yields are uniformly bounded which is not restrictive.
- Assumption ($\mathcal{H}21$) sets a condition on the weightings $(w_j^M)_{j=1\dots M}$. For example, uniform weights do satisfy this condition :

$$\sqrt{\sum_{j=1}^{M}\frac{1}{M^2}} = \frac{1}{\sqrt{M}} \underset{M\to\infty}{\longrightarrow} 0$$

  In Table 4.1, we compute the quantity $(P_w^M)^2$ for the Eurostoxx index and find that it is indeed very small (of the order $\frac{1}{M}$).

- Assumptions ($\mathcal{H}22$) and ($\mathcal{H}23$) are similar. They express the fact that the distance between $(\beta_j^M)_{j=1\dots M}$ and β and the distance between $(\delta_j^M)_{j=1\dots M}$ and δ tends to 0 when M tends to infinity. More importantly, they give us a means of determining the parameters β and δ :

$$\frac{\sum_{j=1}^M w_j^M |\beta_j^M - \beta|}{\sum_{i=1}^M w_i^M} = \mathbb{E}\,|Y_\beta - \beta| \quad \text{and} \quad \frac{\sum_{j=1}^M w_j^M |\delta_j^M - \delta|}{\sum_{i=1}^M w_i^M} = \mathbb{E}\,|Y_\delta - \delta|$$

where Y_β and Y_δ are discrete random variables having the following probability distributions:

$$\forall j \in \{1,\dots,M\}, \qquad \mathbb{P}\left(Y_\beta = \beta_j\right) = \frac{w_j^M}{\sum_{i=1}^M w_i^M} \quad \text{and} \quad \mathbb{P}\left(Y_\delta = \delta_j^M\right) = \frac{w_j^M}{\sum_{i=1}^M w_i^M}.$$

Consequently, the optimal choice of the parameters is the median [4] of Y_β for β and the median of Y_δ for δ. Nevertheless, one does not actually have the choice for the coefficient β. Indeed, recall that by definition of the beta coefficients :

$$\beta_j^M := \frac{Cov(r_j, r_I)}{Var(r_I)} = \frac{\beta_j \beta \sigma^2}{\beta^2 \sigma^2} = \frac{\beta_j}{\beta},$$

so one should take $\beta = 1$. In Table 4.1, we see that the optimal choice of β is very close to 1 and that the quantities of interest, $(P^M_{\beta_{opt}})^2$ and $(P^M_{\beta=1})^2$ are also very close to each other.

$(P_w^M)^2$	β_{opt}	$(P^M_{\beta_{opt}})^2$	$(P^M_{\beta=1})^2$
0.026	0.975	0.0173	0.0174

Table 4.1: Computation of $(P_w^M)^2, \beta_{opt}$ and $(P^M_{\beta_{opt}})^2$ for the Eurostoxx index at December 21, 2007. The beta coefficients are estimated on a two year history.

Simplified model

To sum up, we have shown that, under mild assumptions, when the number of underlying stocks is large, the original model may be approximated by the following dynamics

$$\begin{aligned} \forall j \in \{1,\dots,M\}, \quad \frac{dS_t^j}{S_t^j} &= (r - \delta_j)dt + \beta_j\, \sigma(t, I_t) dB_t + \eta_j(t, S_t^j) dW_t^j \\ \frac{dI_t}{I_t} &= (r - \delta_I)dt + \sigma(t, I_t) dB_t. \end{aligned} \tag{4.6}$$

Interestingly, we end up with a local volatility model for the index and, for each stock, a stochastic volatility model decomposed into a systemic part driven by the index level and an intrinsic part. Note that this simplified model is not valid for options written on the index together

4. The median of a real random variable X is any real number m satisfying :

$$\mathbb{P}(X \le m) \ge \frac{1}{2} \quad \text{and} \quad \mathbb{P}(X \ge m) \ge \frac{1}{2}.$$

It has the property of minimizing the L^1-distance to X : $m = \arg\min_{x \in \mathbb{R}} \mathbb{E}|X - x|$.

with all its composing stocks since the index is no longer an exact, but an approximate, weighted sum of the stocks. In this case, one should consider the reconstructed index $\overline{I}_t^M = \sum_{j=1}^M w_j S_t^j$ or use the original model.

The fact remains that the simplified model can be used for options written on the stocks or on the index or even on the index together with few stocks.

4.3 Model calibration

Calibration, which is how to determine the model parameters in order to fit market prices at best, is of paramount importance in practice. In the following, we try to tackle this issue for both our simplified and original model :

4.3.1 Simplified model

$$\begin{aligned} \forall j \in \{1, \dots, M\}, \quad \frac{dS_t^j}{S_t^j} &= (r - \delta_j)dt + \beta_j\, \sigma(t, I_t)dB_t + \eta_j(t, S_t^j)dW_t^j \\ \frac{dI_t}{I_t} &= (r - \delta_I)dt + \sigma(t, I_t)dB_t \end{aligned} \tag{4.7}$$

The short interest rate and the dividend yields can be extracted from the market. The calibration of the local volatility σ to fit index option prices is a classic problem. What seems to be the market practice is to do a best-fit of a chosen parametric form and match it to the available market prices. This is an important feature of our model : even though the index is reconstructed from the stocks, its calibration remains comparatively easy. Actually our model gives an advantage to the fit of index option prices in comparison with options written on the stocks, which is in line with the market since index options are usually very liquid in comparison with individual stock options.

The calibration of the beta coefficients is more tedious. Indeed, estimation based on historical data can be unsuitable for our model when the historical beta is much larger than the implied one: in this case, since the slope of the local volatility of the index is usually steeper than the one of the stock, the systemic part of the volatility of the stock in our model can be larger than the local volatility of the stock.

To be specific, thanks to the usual formula relating the stochastic volatility to the local volatility (for the theoretical result, see the paper of Gyöngy [53]), one can express the local variance of the stock as

$$v_{loc}(t, K) = \eta^2(t, K) + \beta^2 \mathbb{E}\left(\sigma^2(t, I_t) \,|\, S_t = K\right). \tag{4.8}$$

We see that when $\beta_{hist}^2 \mathbb{E}\left(\sigma^2(t, I_t) \,|\, S_t = K\right)$ becomes larger than $v_{loc}(t, K)$, the local volatility given by our model is larger than the true local volatility of the stock. The right way to handle the estimation of the beta coefficient is then to compute an implied beta calibrated to the options market. Unfortunately, there is no option product that permits us to do this reasonably [5] and one should take a beta coefficient lower than the historical beta whenever the preceding problem is encountered and a beta coefficient higher than the historical one whenever it is possible, such that the following rule of thumb is observed :

5. One financial product that can lead to an easy calibration of the beta coefficient should revolve around the correlation between an index and one of its composing stocks. This is not the case for the most liquid correlation swaps which are sensitive to an average correlation between all the stocks.

$$\sum_{j=1}^{M} w_j \beta_j \simeq 1.$$

In Figure 4.1, we have plotted both the local volatility of the stock, the local volatility of the index, the systemic part of the volatility of the stock $\beta_{hist}\sigma(T, I_T)$ and $\beta_{hist}\mathbb{E}\left(\sigma(T, I_T)|S_T = K\right)$ when η is set to zero (which intuitively gives the lowest local volatility that one can obtain in our model) for a maturity $T = 1$ year. We considered three components of the Eurostoxx : AXA, ALCATEL and CARREFOUR at December 21, 2007. We made this choice deliberately in order to point out the extreme situations one can face :

- AXA is an example of a stock with a high beta coefficient ($\beta = 1.4$).
- CARREFOUR is an example of a stock with a low beta coefficient ($\beta = 0.7$).
- ALCATEL is an example of a stock with a high volatility level but with a low smile effect ($\beta = 1.1$).

The local volatilities are obtained from a parametric fonction of the forward moneyness achieving a best-fit to market smile data. The x-axis represents the moneyness, that is the strike over the spot ($\frac{K}{S_0}$ for a the stock and $\frac{K}{I_0}$ for the index). Clearly, we can deduce that the market is choosing a *beta* coefficient for both AXA and ALCATEL that is lower than the historical one whereas, for CARREFOUR, one can plug the historical beta, or even a larger one, in (4.7) and still be able to calibrate the model.

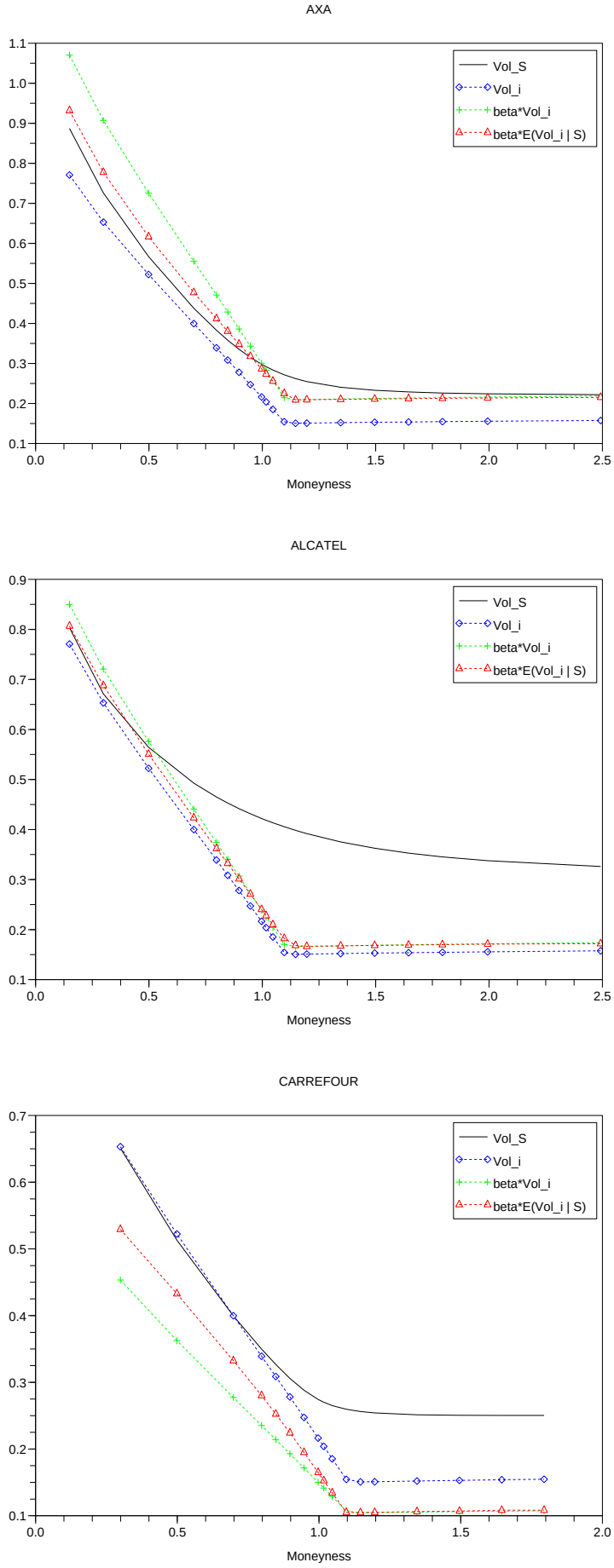

Figure 4.1: Local volatilities of AXA, ALCATEL and CARREFOUR together with $\sigma(T, I_T^{Eurostoxx})$, $\beta_{hist}\sigma(T, I_T^{Eurostoxx})$ and $\beta_{hist}\mathbb{E}\left(\sigma(T, I_T^{Eurostoxx})|S_T = K\right)$ when η is set to zero.

Finally, the remaining parameters that have to be calibrated to fit option prices are the volatility coefficients $\eta_1, \ldots, \eta_M$. From now on, we omit the index j to simplify the notations and we consider the issue of calibrating the volatility coefficient η for a given stock.

From equation (4.8), one gets

$$\eta(t,K) = \sqrt{v_{loc}(t,K) - \beta^2 \mathbb{E}\left(\sigma^2(t,I_t) \,|\, S_t = K\right)}. \tag{4.9}$$

As previously mentioned, v_{loc} can be determined with the best-fit of a parametric form to the stock market smile but determining the conditional expectation is a more challenging task. Note that, since the law of (S_t, I_t) depends on η, so does the conditional expectation and therefore it is difficult to get an estimation of it or to simulate a stochastic differential equation that gives the same vanilla prices as those given by the market. In order to address this issue, we suggest two different simulation based approaches. The first one is based on non-parametric estimation of the conditional expectation and the second one on parametric estimation.

Estimation of the conditional expectation

The idea behind the following techniques is to circumvent the difficulty of calibrating the volatility coefficient η. Indeed, if we plug the formula (4.8) in the dynamics of the stock, we obtain a stochastic differential equation that is nonlinear in the sense of McKean :

$$\begin{aligned} \frac{dS_t}{S_t} &= (r-\delta)dt + \beta\,\sigma(t,I_t)dB_t + \sqrt{v_{loc}(t,S_t) - \beta^2 \mathbb{E}\left(\sigma^2(t,I_t) \,|\, S_t\right)}dW_t \\ \frac{dI_t}{I_t} &= (r-\delta_I)dt + \sigma(t,I_t)dB_t \end{aligned} \tag{4.10}$$

For an introduction to the topics of nonlinear stochastic differential equations and propagation of chaos, we refer to the lecture notes of Sznitman [116] and Méléard [89]. In our case, the nonlinearity appears in the diffusion coefficient through the conditional expectation term. This makes the natural question of existence and uniqueness of a solution very difficult to handle. The case of a drift involving a conditional expectation has only been handled recently even for constant diffusion coefficient (see Talay and Vaillant [118] and Dermoune [31]). Meanwhile, it is possible to simulate such a stochastic differential equation by means of a system of N interacting paths using either a non-parametric estimation of the conditional expectation or regression techniques. The advantage of the regression approach over the non-parametric estimation is that it also yields a smooth approximation of the function $\mathbb{E}\left(\sigma^2(t,I_t) \,|\, S_t = s\right)$ whereas, with a non-parametric method, one has to interpolate the estimated function and to carefully tune the window parameter to obtain a smooth approximation.

Non-parametric estimation

Non-parametric estimators of the conditional expectation, and more generally non-parametric density estimators, have been widely studied in the literature. We will focus on kernel estimators of the Nadaraya-Watson type (see Watson [130] and Nadaraya [93]) : given N observations $(S_t^i, I_t^i)_{i=1\ldots N}$ of (S_t, I_t), we consider the kernel conditional expectation estimator of $\mathbb{E}\left(\sigma^2(t,I_t) \,|\, S_t = s\right)$ given by

$$\frac{\sum_{i=1}^{N} \sigma^2(t, I_t^i) K\left(\frac{s - S_t^i}{h_N}\right)}{\sum_{i=1}^{N} K\left(\frac{s - S_t^i}{h_N}\right)}$$

where K is a non-negative kernel such that $\int_{\mathbb{R}} K(x)dx = 1$ and h_N is a smoothing parameter which tends to zero as $N \to +\infty$. This leads to the following system with N interacting particles : $\forall\, 1 \leq i \leq N$,

$$\begin{cases} \frac{dS_t^{i,N}}{S_t^{i,N}} = (r-\delta)dt + \beta\,\sigma(t, I_t^i)dB_t^i + \sqrt{v_{loc}(t, S_t^{i,N}) - \beta^2 \frac{\sum_{j=1}^N \sigma^2(t, I_t^j) K\left(\frac{S_t^{i,N} - S_t^{j,N}}{h_N}\right)}{\sum_{j=1}^N K\left(\frac{S_t^{i,N} - S_t^{j,N}}{h_N}\right)}} dW_t^i,\ S_0^{i,N} = S_0 \\ \frac{dI_t^i}{I_t^i} = (r - \delta_I)dt + \sigma(t, I_t^i)dB_t^i,\ I_0^i = I_0 \end{cases}$$

where $(B^i, W^i)_{i\geq 1}$ is a sequence of independent two-dimensional Brownian motions. This $2N$-dimensional SDE may be discretized using the Euler scheme :

Let $0 = t_0 < \cdots < t_M = T$ be a subdivision with step $\frac{T}{M}$ of $[0, T]$. For each $k \in \{0, \ldots, M-1\}$, $\forall\, 1 \leq i \leq N$,

$$\begin{cases} \overline{S}_{t_{k+1}}^{i,N} = \overline{S}_{t_k}^{i,N} \left((r-\delta)\frac{T}{M} + \beta\,\sigma(t_k, \overline{I}_{t_k}^i)\sqrt{\frac{T}{M}} G_{i,k}^1 + \sqrt{v_{loc}(t_k, \overline{S}_{t_k}^{i,N}) - \beta^2 \frac{\sum_{j=1}^N \sigma^2(t_k, \overline{I}_{t_k}^j) K\left(\frac{\overline{S}_{t_k}^{i,N} - \overline{S}_{t_k}^{j,N}}{h_N}\right)}{\sum_{j=1}^N K\left(\frac{\overline{S}_{t_k}^{i,N} - \overline{S}_{t_k}^{j,N}}{h_N}\right)}} \sqrt{\frac{T}{M}} G_{i,k}^2 \right) \\ \overline{I}_{t_{k+1}}^i = \overline{I}_{t_k}^i \left((r - \delta_I)\frac{T}{M} + \sigma(t_k, \overline{I}_{t_k}^i)\sqrt{\frac{T}{M}} G_{i,k}^1 \right) \end{cases}$$

where $(G_{i,k}^1)_{1\leq i\leq N, 0\leq k\leq M-1}$ and $(G_{i,k}^2)_{1\leq i\leq N, 0\leq k\leq M-1}$ are independent centered and reduced Gaussian random variables.

<u>Parametric estimation</u>

Another approach to estimate conditional expectations is to use parametric estimators, or projection. This idea has also been widely used and studied previously (for example in finance, one can think of the Longstaff-Schwartz algorithm for pricing American options Longstaff and Schwartz [84]). Noting that the conditional expectation is a projection operator on the space of square integrable random variables, one can approximate $\mathbb{E}\left(\sigma^2(t, I_t) \,|\, S_t = s\right)$ by the parametric estimator

$$\sum_{k=1}^{K} \alpha_k f_k(s)$$

where $(f_k)_{k=1\ldots K}$ is a functional basis and $\alpha = (\alpha_k)_{k=1\ldots K}$ is a vector of parameters estimated by least mean squares : given N observations $(S_t^i, I_t^i)_{i=1\ldots N}$ of (S_t, I_t), α minimizes $\sum_{i=1}^N \left(\sigma^2(t, I_t^i) - \sum_{k=1}^K \alpha_k f_k(S_t^i)\right)^2$.

Numerical results

A toy model

In the first numerical example, we suppose that the local volatility of the stock is constant and we try to reconstruct it by simulating the particle system of the non-parametric method presented above. We consider the Eurostoxx index and we determine its local volatility by fitting the market prices at December 21, 2007.

As described above, we can approximate the following SDE using a system of N interacting particles :

$$\begin{aligned} \frac{dS_t}{S_t} &= (r-\delta)dt + \beta\,\sigma(t,I_t)dB_t + \sqrt{v - \beta^2 \mathbb{E}\left(\sigma^2(t,I_t) \mid S_t\right)}dW_t \\ \frac{dI_t}{I_t} &= (r-\delta_I)dt + \sigma(t,I_t)dB_t \end{aligned} \tag{4.11}$$

Using these simulations to price European call options for different strikes, one should obtain the same results as a Black & Scholes model with volatility $\sqrt{v}$. In Figure 4.2, we plot the implied volatility obtained by independent simulations of $N = 5000$ paths and see that the implied volatilities obtained are indeed close to the exact volatility level. This example was generated with the following arbitrary set of parameters :

- $S_0 = 100$.
- $\beta = 0.7$.
- $r = 0.05$.
- $\delta = \delta_I = 0$.
- $\sqrt{v} = 0.3$.
- $T = 1$.
- Number of simulated paths : $N = 5000$.
- Number of time steps in the Euler scheme : $M = 20$.

In this example and for all the following numerical experiments, we use a Gaussian kernel : $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$. The smoothing parameter h_N is set to $N^{-\frac{1}{5}}$ which is the optimal bandwidth that one obtains when minimizing the asymptotic mean square error of the Nadaraya-Watson estimator under some regularity assumptions and assuming independence of the random variables involved (see for example Bosq [17]).

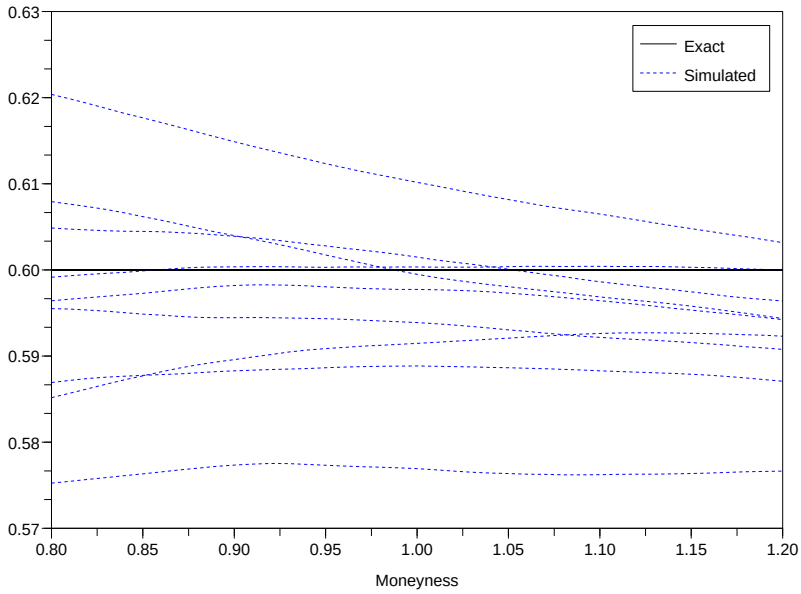

Figure 4.2: Implied volatility obtained for nine independent simulations with $N = 5000$ paths.

An example with real data

In the following, we test our model with real data. More precisely, given the local volatilities of the Eurostoxx index and of Carrefour at December 21, 2007, we simulate the particle system (4.10) by different methods for a one year maturity.

1. An acceleration technique

   The simulation of the particle system is very time consuming : for each discretization step and for each stock particle, one has to make N computations which yield a global complexity of order $O(MN^2)$ where M is the number of time steps in the Euler scheme. Acceleration techniques are thus unavoidable. One possible method consists in reducing the number of interactions : instead of making N computations for each estimation of the conditional expectation, one can neglect interactions which involve particles which are far away from each other. When the kernel used is non increasing with the absolute value of its argument, the easiest way to implement this idea is to sort the particles at each step and, whenever a contribution of a particle is lower than some fixed threshold, to stop the estimation of the conditional expectation.

   Of course, by doing this, we lose in precision for the same number of interacting particles, especially for deep in/out of the money strikes. But what we gain in terms of computation time is much more important : in Figure 4.3, we plot the implied volatility obtained by the naive method and the method with the above acceleration technique for the same number

$N = 10000$ of particles. We take as threshold $\frac{1}{N}$ and set $h_N = N^{-\frac{1}{10}}$ for the bandwidth parameter [6] and $M = 20$ for the number of time steps in the Euler scheme. The computation time, on a computer with a 2.8 Ghz Intel Penthium 4 processor, is of 52 minutes for the naive method and of 5 minutes for the accelerated one.

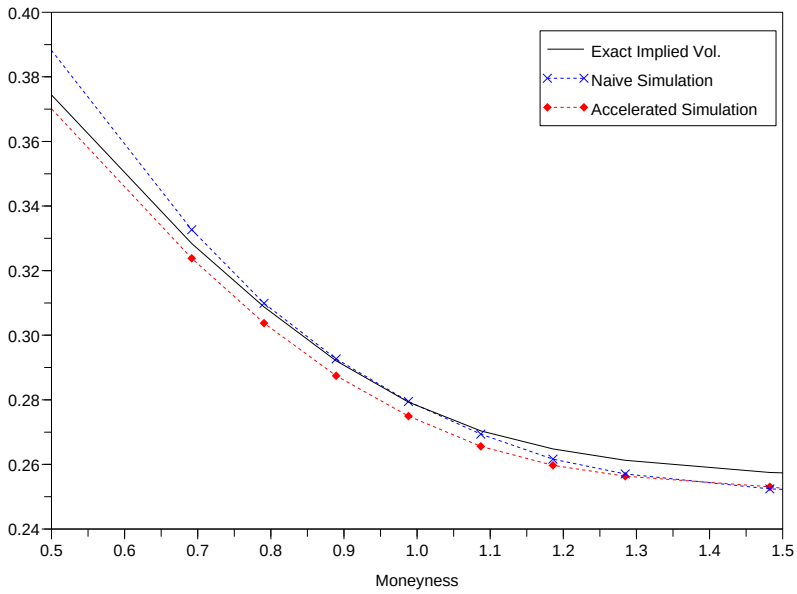

Figure 4.3: Comparison between the naive technique and the accelerated one.

More importantly, we see that the implied volatility $\widehat{\sigma}_{simul}$ obtained by simulations converges to the exact volatility $\widehat{\sigma}_{exact}$: see Figure 4.4 and Table 4.2. With a reasonable number of simulated paths, $N = 200000$, the error on the implied volatility remains clearly tolerable for practitioners (of the order of 10 bp) except for a deep in the money call ($K = 0.3S_0$) where it attains 195 bp.

Moneyness ($\frac{K}{S_0}$)	0.30	0.49	0.69	0.79	0.89	0.99	1.09	1.19	1.28	1.48	1.98
Error : $\lvert\widehat{\sigma}_{simul} - \widehat{\sigma}_{exact}\rvert$	195	36	8	5	2	1	2	9	17	32	56

Table 4.2: Error (in bp) on the implied volatility with $N = 200000$ particles.

6. In order to smooth the estimation, one has to choose a bandwidth parameter that is greater than the theoretical optimal parameter $N^{-\frac{1}{5}}$.

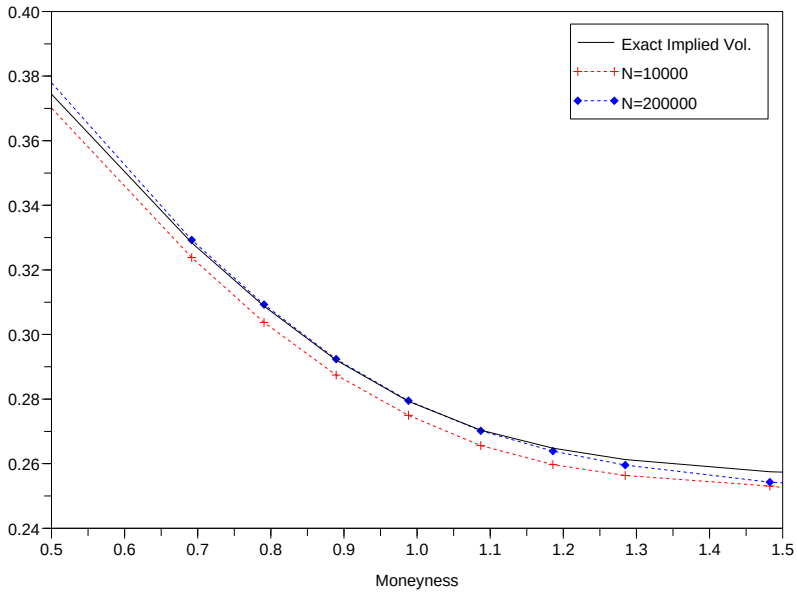

Figure 4.4: Convergence of the implied volatility obtained with non-parametric estimation.

2. Independent particles

   Unlike the parametric method, non-parametric estimation of the conditional expectation gives the value of the intrinsic volatility η at the simulated points only. However, using an interpolation technique, one can first reconstruct η with N_1 dependent particles and then simulate the 2-dimensional stochastic differential equation with N_2 independent draws, N_2 being larger than N_1. By doing so, we speed up the simulations but one has to choose carefully the size N_1 of the particle system in order to have a reasonable estimation of the intrinsic volatility and to tune the bandwidth parameter in order to smooth the estimation (our numerical tests were done with $N_1 = 1000, N_2 = 100000$ and $h_{N_1} = N_1^{-\frac{1}{10}}$). In Figures 4.5 and 4.6, we give the surfaces of both the local volatility and the intrinsic volatility of the stock. This latter is used to draw independent simulations of the index along with the stock and we see in Figure 4.7 that the implied volatility obtained is close to the right one, especially near the money.

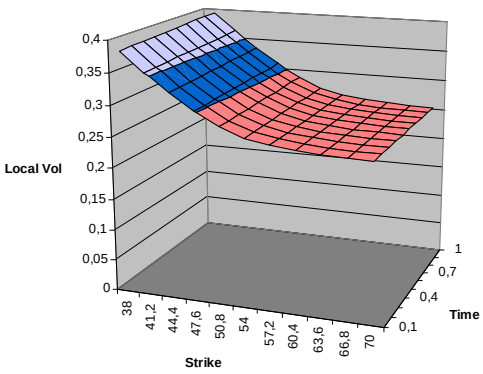

Figure 4.5: Local volatility surface of the stock.

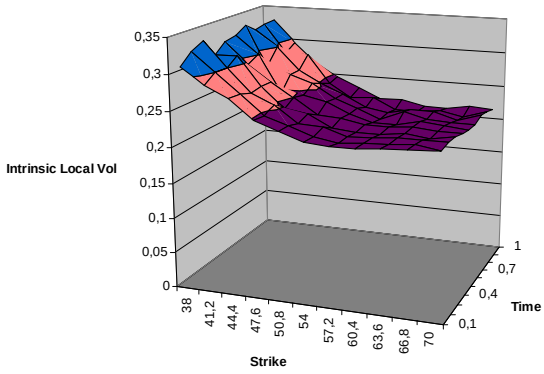

Figure 4.6: Intrinsic part of the stochastic volatility.

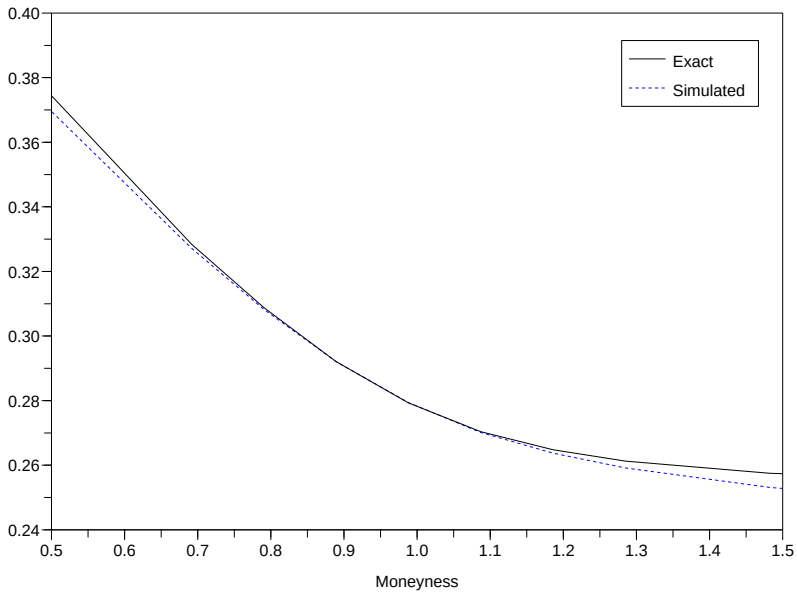

Figure 4.7: Simulated implied volatility with independent draws.

4.3.2 Original model

We now turn to the calibration of our original model :

$$\forall j \in \{1, \dots, M\}, \quad \frac{dS_t^{j,M}}{S_t^{j,M}} = (r - \delta_j)dt + \beta_j\, \sigma(t, I_t^M)dB_t + \eta_j(t, S_t^{j,M})dW_t^j \tag{4.12}$$

with $I_t^M = \sum_{i=1}^M w_i S_t^{i,M}$.

Obviously, it is rather complicated to have a perfect calibration for both index and stocks within this framework. Nevertheless, one can either

- take for σ the calibrated local volatility of the index and then calibrate the volatility coefficients η_j using an adaptation of the non-parametric method presented above in order to fit all the individual stock smiles at the same time. In this case, the index is not perfectly calibrated but, thanks to Theorem 35, one can expect the error to be small.

Or,

- take for σ and η_j the calibrated coefficients in the simplified model framework. Once again, the calibration is not perfect and this time for both index and individual stocks but Theorems 35 and 36 suggest that the calibration error will be negligible.

Hence, in comparison with the simplified model, we allow ourselves a slight error in the calibration but we guarantee the additivity constraint $I_t^M = \sum_{i=1}^M w_i S_t^{i,M}$. In what follows, we illustrate the effect of Theorems 35 and 36 and compare our models with a constant correlation model.

4.4 Illustration of Theorems 35 and 36 and comparison with a constant correlation model

The objective of this section is to compare index and individual stock smiles obtained with three different models : our original model (4.12), the simplified one (after letting $M \to \infty$) and a model with constant correlation coefficient. More precisely, we consider the following dynamics

1. The original model

$$\begin{aligned} &\forall j \in \{1, \dots, M\}, \quad \frac{dS_t^{j,M}}{S_t^{j,M}} = rdt + \sigma(t, I_t^M)dB_t + \eta(t, S_t^{j,M})dW_t^j \\ &\text{with } I_t^M = \sum_{i=1}^M w_i S_t^{i,M}. \end{aligned} \tag{4.13}$$

2. The simplified model

$$\begin{aligned} \forall j \in \{1, \dots, M\}, \quad \frac{dS_t^j}{S_t^j} &= rdt + \sigma(t, I_t)dB_t + \eta(t, S_t^j)dW_t^j \\ \frac{dI_t}{I_t} &= rdt + \sigma(t, I_t)dB_t. \end{aligned} \tag{4.14}$$

Where we can also compute the reconstructed index $\overline{I}_t^M = \sum_{i=1}^M w_i S_t^i$.

3. The "Market" model

$$\forall j \in \{1, \ldots, M\}, \frac{dS_t^j}{S_t^j} = rdt + \sqrt{v_{loc}(t, S_t^j)} d\widetilde{W}_t^j \tag{4.15}$$

with, $\forall i \neq j$, $d < \widetilde{W}^i, \widetilde{W}^j >_t = \rho\, dt$.

We deliberately dropped the dividend yields and the beta coefficients in order to simplify the numerical experiment. For the volatility coefficient σ, we take as previously the calibrated local volatility of the Eurostoxx. We choose an arbitrary parametric form, fonction of the forward moneyness, for the volatility coefficient η and we evaluate v_{loc} such that the market model and the simplified model yield the same implied volatility for individual stocks. Indeed, it suffices to take

$$v_{loc}(t, s) = \eta^2(t, s) + \mathbb{E}(\sigma^2(t, I_t)|S_t = s)$$

where the conditional expectation can be approximated using the non-parametric method presented above.

Finally, we fix the correlation coefficient ρ such that the market model and the simplified one have the same ATM implied volatility for the index.

The implied volatilities for the index and for an individual stocks obtained by the three models are plotted in Figures 4.8 and 4.9. We also give the difference in basis points between the implied volatilities obtained with the simplified model and the original one in Tables 4.3, 4.4 and 4.5. The parameters we use in our numerical experiment are the following :

- $S_0^1 = \cdots = S_0^M = 53$.
- M, I_0 and the weights $w_1, \ldots, w_M$: the same as of the Eurostoxx index at December 21, 2007.
- $r = 0.045$.
- Maturity $T = 1$ year.
- Number of time steps: 10.
- Number of simulated paths : 100000.

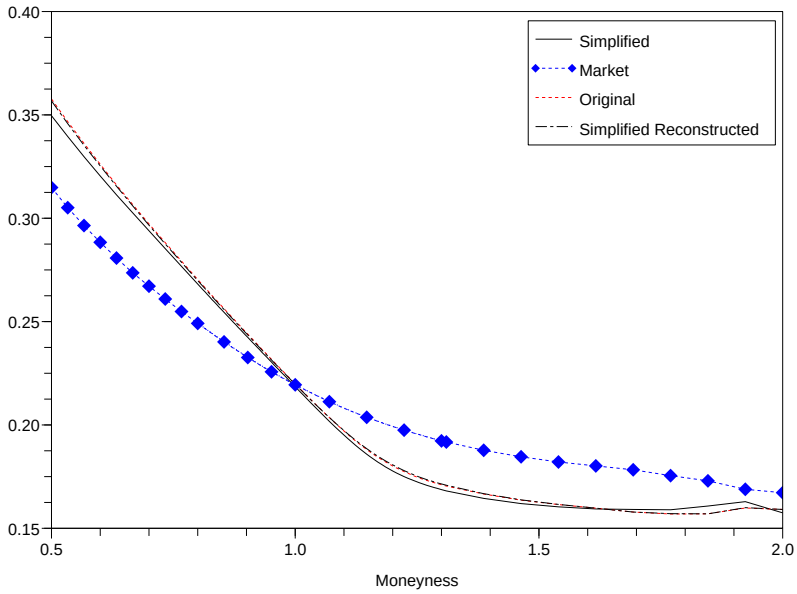

Figure 4.8: Implied volatility of the index.

Moneyness ($\frac{K}{I_0}$)	0.5	0.8	0.9	0.95	1	1.05	1.1	1.2	1.3	1.55	1.85	2
$\lvert\widehat{\sigma}_{simplified} - \widehat{\sigma}_{original}\rvert$	81	22	16	14	14	17	20	24	24	11	38	17

Table 4.3: Difference (in bp) between the index implied volatility obtained with the simplified model and the one obtained with the original model.

Moneyness ($\frac{K}{I_0}$)	0.5	0.8	0.9	0.95	1	1.05	1.1	1.2	1.3	1.55	1.85	2
$\lvert\widehat{\sigma}_{reconstruct} - \widehat{\sigma}_{original}\rvert$	10	5	4	3	2	1	2	5	4	1	0	0

Table 4.4: Difference (in bp) between the implied volatility of the reconstructed index $\overline{I}^M$ in the simplified model and the index implied volatility obtained with the original model.

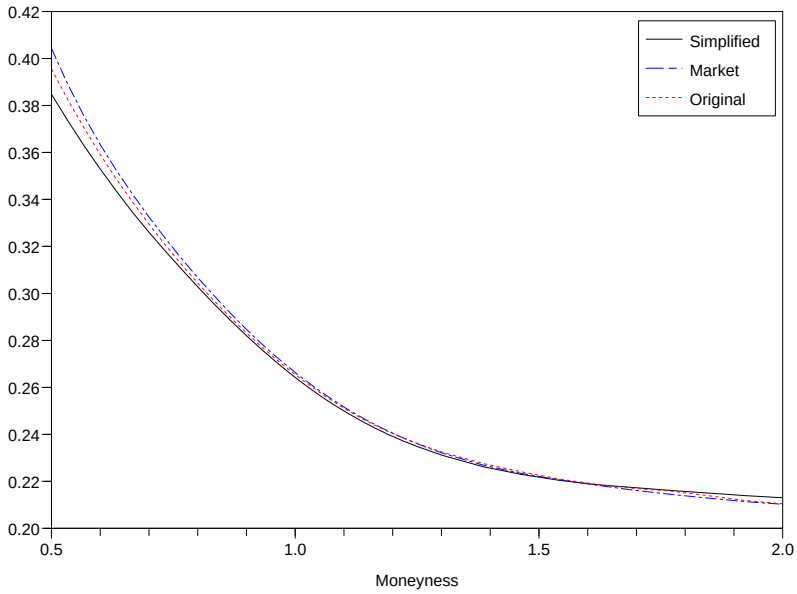

Figure 4.9: Implied volatility of an individual stock.

Moneyness ($\frac{K}{S_0}$)	0.5	0.8	0.9	0.95	1	1.05	1.1	1.2	1.3	1.55	1.85	2
$\lvert\widehat{\sigma}_{simplified} - \widehat{\sigma}_{original}\rvert$	81	22	16	14	14	17	20	24	24	11	38	17

Table 4.5: Difference (in bp) between an individual stock implied volatility obtained with the simplified model and the one obtained with the original model.

As suggested by Theorems 35 and 36, we see that the original model and the simplified one yield implied volatility curves that are very close to each other, both for the index and for individual stocks. The difference in basis points between the implied volatilities is reasonable, especially between the reconstructed index implied volatility of the simplified model and the index implied volatility of the original model.

Concerning the market model, by construction we have the same implied volatility of an individual stock as for the simplified model but the implied volatility of the index obtained is far from the right one, especially the slope of the smile out-of-the-money. This phenomenon is well known in practice (see Bakshi *et al.* [6], Bollen and Whaley [16] or Branger and Schlag [18]) : the implied volatility smile of an index is much steeper than the implied volatility smile of an individual stock, hence the market model of constantly correlated stocks is unable to retrieve the shape of the index smile. More sophisticated dependence structure between stocks is needed. Our modeling framework circumvents this difficulty since we force the index to have the correct volatility smile while the individual stocks can still be properly calibrated.

4.4.1 Application: Pricing of a worst-of option

Apart from handling both the index and its composing stocks, our models are also relevant for the widespread financial products that are sensitive to correlation in the equity world, such as rainbow options.

One example of such products is the worst-of performance option whose payout is referenced to the worst performer in a basket of shares. For a basket of M shares, the payoff of a call with strike K and maturity T writes $\left(\min_{1\leq i\leq M} \frac{S_T^i}{S_0^i} - K\right)_+$. Our objective is to compare the prices obtained by our model to the prices obtained by the market model of constantly correlated stocks. The parameters of the numerical experiment are the same as previously and we set the correlation coefficient ρ such that all the models exhibit the same ATM implied volatility for the index.

The result, as can be seen in Figure 4.10, is that our prices are always lower than the market model price, especially in the money. Hence, a model with constant correlation coefficient, calibrated in order to fit at the money prices, will always overestimate the risks of worst-of options. Note that the prices obtained with the original model and the simplified one are barely distinguishable from each other.

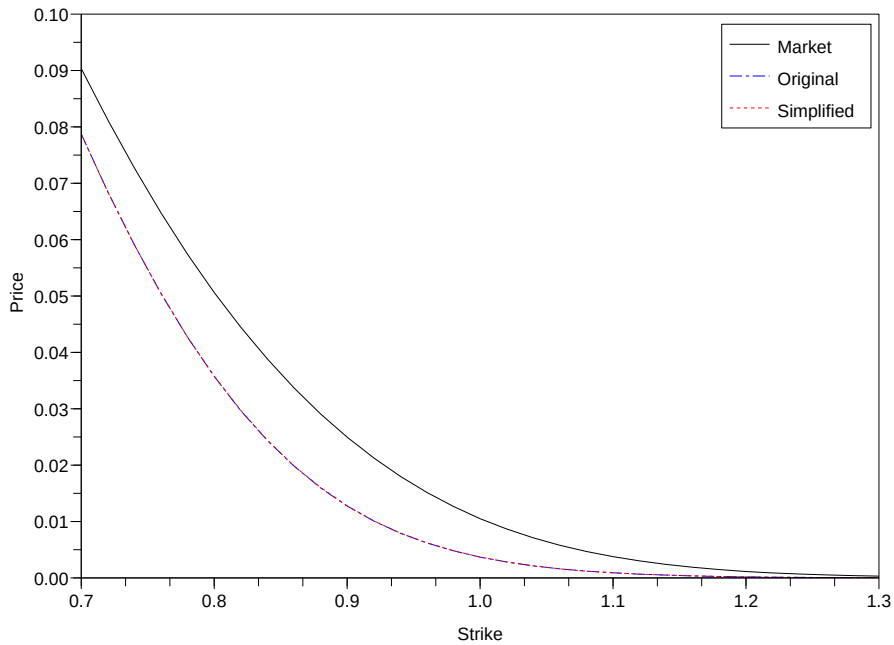

Figure 4.10: Worst-of price.

4.5 Conclusion

In this paper, we have introduced a new model for describing the joint evolution of an index and its composing stocks. The idea behind our view is that an index is not only a weighted sum of stocks but can also be seen as a market factor that influences their dynamics. In order to have a more tractable model, we have studied the limit when the number of underlying stocks goes to infinity and we have shown that our model reduces to a local volatility model for the index and to a stochastic volatility model for each individual stock with volatility driven by the index. Unlike the existing models, we favor the fit of the index smile in comparison with the fit of the stock smiles which goes in accordance with the market since index options are usually more liquid than options on a given stock. We have discussed calibration issues and proposed a simulation-based technique for the calibration of the stock dynamics, which permits us to fit both index and stocks smiles. The numerical results obtained on real data for the Eurostoxx index are very encouraging, especially for accelerated techniques. We have also compared our models (before and after passing to the limit) to a market standard model consisting of local volatility models for the stocks which are constantly correlated and we have seen that it is not possible to retrieve the shape of the index smile. Finally, when considering the pricing of worst-of performance options, which are sensitive to the dependence structure between stocks, we have found that our prices are more aggressive than the prices obtained by the standard market model.

To sum up, we list some properties of our models depending on the options one wishes to handle in the Table below

Purpose	Simplified model	Original model
Options written on -few ($J << M$) stocks -the index.	-Simulation of a $(J+1)$-dimensional SDE : $(I, S^1, \ldots, S^J)$. -Exact calibration of $(S^j)_{1\leq j\leq J}$ and I possible.	-Simulation of an M-dimensional SDE : $(S^{1,M}, \ldots, S^{M,M})$. -Exact calibration of $(S^{j,M})_{1\leq j\leq J}$ possible. -Approximate calibration of I^M.
Options written on -all the stocks -the index.	-Simulation of an $(M+1)$-dimensional SDE : $(I, S^1, \ldots, S^M)$. -Exact calibration of all the stocks possible. -Index value : $\overline{I}_t^M = \sum_{j=1}^M w_j S_t^j$. -Approximate calibration of $\overline{I}^M$.	-Simulation of an M-dimensional SDE : $(S^{1,M}, \ldots, S^{M,M})$. -Exact calibration of all the stocks possible. -Approximate calibration of I^M.

Table 4.6: Which model to use and when.

4.6 Appendix

In order to prove the Theorems 35 and 36, we need the following technical estimation

Lemma 38 — *Under assumption ($\mathcal{H}17$), for all $p \geq 1$, one has*

$$\forall j \in \{1, \ldots, M\}, \quad \sup_{0 \leq t \leq T} \mathbb{E}\left(|S_t^{j,M}|^{2p}\right) \leq C_p \tag{4.16}$$

where $C_p = \max_{1 \leq j \leq M} |S_0^{j,M}|^{2p} \exp\left(\left(2r + (2p-1)(\max_{j \geq 1} \beta_j^2 + 1)K_b^2\right) pT\right)$.

Proof : By Itô's lemma one has

$$\begin{aligned} |S_t^{j,M}|^{2p} &= |S_0^{j,M}|^{2p} + \int_0^t |S_s^{j,M}|^{2p}((2p)(r-\delta_j) + p(2p-1)(\beta_j^2\sigma^2(s, I_s^M) + \eta_j^2(s, S_s^{j,M})))ds \\ &\quad + \int_0^t (2p)|S_s^{j,M}|^{2p}(\beta_j\sigma(s, I_s^M)dB_s + \eta_j(s, S_s^{j,M})dW_s^j) \end{aligned}$$

In order to get rid of the stochastic integral, we use a localization technique : let ν_n be the stopping time defined for each $n \in \mathbb{N}$ by $\nu_n := \inf\{t \geq 0; |S_t^{j,M}| \geq n\}$. Then, using ($\mathcal{H}17$), one has

$$\begin{aligned} \mathbb{E}\left(|S_{t\wedge\nu_n}^{j,M}|^{2p}\right) &= |S_0^{j,M}|^{2p} + \mathbb{E}\left(\int_0^{t\wedge\nu_n} |S_s^{j,M}|^{2p}((2p)(r-\delta_j) + p(2p-1)(\beta_j^2\sigma^2(s, I_s^M) + \eta_j^2(s, S_s^{j,M}))ds\right) \\ &\leq |S_0^{j,M}|^{2p} + \left((2p)(r-\delta_j)\mathbb{1}_{\{r-\delta_j \geq 0\}} + p(2p-1)(\beta_j^2+1)K_b^2\right) \int_0^t \mathbb{E}\left(|S_{s\wedge\nu_n}^{j,M}|^{2p}\right) ds \end{aligned}$$

So, by Gronwall's lemma and the fact that the dividends are nonnegative,

$$\forall t \leq T, \mathbb{E}\left(|S_{t\wedge\nu_n}^{j,M}|^{2p}\right) \leq |S_0^{j,M}|^{2p} \exp\left(\left(2rp + p(2p-1)(\beta_j^2+1)K_b^2\right)T\right) \tag{4.17}$$

Finally, Fatou's lemma permits us to conclude :

$$\sup_{0 \leq t \leq T} \mathbb{E}\left(|S_t^{j,M}|^{2p}\right) \leq |S_0^{j,M}|^{2p} \exp\left(\left(2rp + p(2p-1)(\beta_j^2+1)K_b^2\right)T\right). \tag{4.18}$$

□

4.6.1 Proof of Theorem 35

Using the SDEs (4.3) and (4.5), one has

$$
\begin{aligned}
|I_t^M - I_t|^{2p} &= \Big| r\int_0^t \left(I_s^M - I_s\right) ds - \int_0^t \left(\sum_{j=1}^M \delta_j w_j S_s^{j,M} - \delta I_s\right) ds \\
&\quad + \int_0^t \left(\sum_{j=1}^M \beta_j w_j S_s^{j,M}\sigma(s, I_s^M) - \beta I_s \sigma(s, I_s)\right) dB_s + \sum_{j=1}^M w_j \int_0^t S_s^{j,M}\eta_j(s, S_s^{j,M}) dW_s^j \Big|^{2p} \\
&\leq 4^{2p-1}\Bigg(r^{2p} t^{2p-1} \int_0^t (I_s^M - I_s)^{2p} ds + t^{2p-1}\int_0^t \left(\sum_{j=1}^M \delta_j w_j S_s^{j,M} - \delta I_s\right)^{2p} ds \\
&\quad + \Big|\int_0^t \left(\sum_{j=1}^M \beta_j w_j S_s^{j,M}\sigma(s, I_s^M) - \beta I_s \sigma(s, I_s)\right) dB_s\Big|^{2p} + \Big|\sum_{j=1}^M w_j \int_0^t S_s^{j,M}\eta_j(s, S_s^{j,M}) dW_s^j\Big|^{2p}\Bigg)
\end{aligned}
$$

Hence, using the Burkholder-Davis-Gundy inequality (see Karatzas and Shreve [63] p. 166), there exists a universal positive constant K_p such that

$$
\mathbb{E}\left(\sup_{0\leq t\leq T} |I_t^M - I_t|^{2p}\right) \leq 4^{2p-1}(a_M + b_M + c_M + d_M)
$$

where

- $a_M = r^{2p}\, T^{2p-1} \int_0^T \mathbb{E}\big((I_s^M - I_s)^{2p}\big) ds$
- $b_M = T^{2p-1} \int_0^T \mathbb{E}\left(\left(\sum_{j=1}^M \delta_j w_j S_s^{j,M} - \delta I_s\right)^{2p}\right) ds$
- $c_M = K_p T^{p-1} \int_0^T \mathbb{E}\left(\left(\sum_{j=1}^M \beta_j w_j S_s^{j,M}\sigma(s, I_s^M) - \beta I_s \sigma(s, I_s)\right)^{2p}\right) ds$
- $d_M = K_p T^{p-1} \int_0^T \mathbb{E}\left(\left(\sum_{j=1}^M \left(w_j S_s^{j,M}\eta_j(s, S_s^{j,M})\right)^2\right)^p\right) ds$

The term a_M is the easiest one to handle :

$$
a_M \leq r^{2p}\, T^{2p-1} \int_0^T \mathbb{E}\left(\sup_{0\leq u\leq s} |I_u^M - I_u|^{2p}\right) ds. \tag{4.19}
$$

Next, using assumption ($\mathcal{H}17$) for the first inequality, Hölder's inequality for the second and lemma 38 for the third, one gets

$$\begin{aligned}
d_M &= K_p T^{p-1} \int_0^T \sum_{j_1=1}^M \cdots \sum_{j_p=1}^M \mathbb{E}\left(\prod_{k=1}^p w_{j_k}^2 (S_s^{j_k,M})^2 (\eta_{j_k}(s, S_s^{j_k,M}))^2\right) ds \\
&\leq K_p K_b^{2p} T^{p-1} \int_0^T \sum_{j_1=1}^M \cdots \sum_{j_p=1}^M (\prod_{k=1}^p w_{j_k}^2) \mathbb{E}\left(\prod_{k=1}^p (S_s^{j_k,M})^2\right) ds \\
&\leq K_p K_b^{2p} T^{p-1} \int_0^T \sum_{j_1=1}^M \cdots \sum_{j_p=1}^M \prod_{k=1}^p w_{j_k}^2 \left(\mathbb{E}\left((S_s^{j_k,M})^{2p}\right)\right)^{\frac{1}{p}} ds \\
&\leq K_p K_b^{2p} T^p C_p \left(\sum_{j=1}^M w_j^2\right)^p
\end{aligned} \tag{4.20}$$

The same arguments enable us to control the term b_M :

$$\begin{aligned}
b_M &= T^{2p-1} \int_0^T \mathbb{E}\left(\left(\sum_{j=1}^M \delta_j w_j S_s^{j,M} - \delta I_s\right)^{2p}\right) ds \\
&\leq (2T)^{2p-1} \left(\int_0^T \mathbb{E}\left(\left(\sum_{j=1}^M \delta_j w_j S_s^{j,M} - \delta I_s^M\right)^{2p}\right) + \mathbb{E}\left(\left(\delta I_s^M - \delta I_s\right)^{2p}\right) ds\right) \\
&\leq (2T)^{2p-1} \int_0^T \mathbb{E}\left(\left(\sum_{j=1}^M (\delta_j - \delta) w_j S_s^{j,M}\right)^{2p}\right) ds + (2T)^{2p-1} \delta^{2p} \int_0^T \mathbb{E}\left(\sup_{0\leq u\leq s} |I_u^M - I_u|^{2p}\right) ds \\
&\leq 2^{2p-1} T^{2p} C_p \left(\sum_{j=1}^M w_j |\delta_j - \delta|\right)^{2p} + (2T)^{2p-1} \delta^{2p} \int_0^T \mathbb{E}\left(\sup_{0\leq u\leq s} |I_u^M - I_u|^{2p}\right) ds.
\end{aligned} \tag{4.21}$$

For the remaining term c_M, we will also need the Lipschitz assumption ($\mathcal{H}18$)

$$\begin{aligned}
c_M &= K_p T^{p-1} \int_0^T \mathbb{E}\left(\left(\sum_{j=1}^M \beta_j w_j S_s^{j,M} \sigma(s, I_s^M) - \beta I_s \sigma(s, I_s)\right)^{2p}\right) ds \\
&\leq 2^{2p-1} K_p T^{p-1} \left(\int_0^T \mathbb{E}\left(\left(\sum_{j=1}^M (\beta_j - \beta) w_j S_s^{j,M} \sigma(s, I_s^M)\right)^{2p}\right) + \mathbb{E}\left((\beta I_s^M \sigma(s, I_s^M) - \beta I_s \sigma(s, I_s))^{2p}\right) ds\right) \\
&\leq 2^{2p-1} K_p T^p K_b^{2p} C_p \left(\sum_{j=1}^M w_j |\beta_j - \beta|\right)^{2p} + 2^{2p-1} K_p T^{p-1} (\beta K_\sigma)^{2p} \int_0^T \mathbb{E}\left(\sup_{0\leq u\leq s} |I_u^M - I_u|^{2p}\right) ds.
\end{aligned} \tag{4.22}$$

So, combining the inequalities (4.19), (4.20), (4.21) and (4.22), one obtains

$$\begin{aligned}\mathbb{E}\left(\sup_{0\le t\le T}|I_t^M-I_t|^{2p}\right) &\le C_0\left(\left(\sum_{j=1}^M w_j^2\right)^p+\left(\sum_{j=1}^M w_j|\beta_j-\beta|\right)^{2p}+\left(\sum_{j=1}^M w_j|\delta_j-\delta|\right)^{2p}\right)\\ &\quad +C_1\int_0^T\mathbb{E}\left(\sup_{0\le u\le s}|I_u^M-I_u|^2\right)ds\end{aligned}$$

with $C_0=8^{2p-1}T^p(T^p+K_pK_b^{2p})C_p$ and $C_1=4^{2p-1}(2^{2p-1}K_pT^{p-1}(\beta K_\sigma)^{2p}+(2T)^{2p-1}\delta^{2p}+r^{2p}\,T^{2p-1})$.

Finally, by means of Gronwall's lemma, we conclude that

$$\mathbb{E}\left(\sup_{0\le t\le T}|I_t^M-I_t|^{2p}\right)\le C_T\left(\left(\sum_{j=1}^M w_j^2\right)^p+\left(\sum_{j=1}^M w_j|\beta_j-\beta|\right)^{2p}+\left(\sum_{j=1}^M w_j|\delta_j-\delta|\right)^{2p}\right)$$

where

$$C_T=C_0e^{C_1T}.$$

4.6.2 Proof of Theorem 36

The proof is similar to the previous one :

$$\begin{aligned}|S_t^{j,M}-S_t^j|^{2p} &\le 3^{2p-1}\left((r-\delta_j)^{2p}t^{2p-1}\int_0^t(S_s^{j,M}-S_s^j)^{2p}ds+\left|\int_0^t(S_s^{j,M}\eta_j(s,S_s^{j,M})-S_s^j\eta_j(s,S_s^j))dW_s^j\right|^{2p}\right.\\ &\quad\left.+\beta_j^{2p}\left|\int_0^t(S_s^{j,M}\sigma(s,I_s^M)-S_s^j\sigma(s,I_s))dB_s\right|^{2p}\right)\end{aligned}$$

hence, using the Burkholder-Davis-Gundy inequality, there exists a constant K_p such that

$$\begin{aligned}\mathbb{E}\left(\sup_{0\le t\le T}|S_t^{j,M}-S_t^j|^{2p}\right) &\le 3^{2p-1}\left((r-\delta_j)^{2p}T^{2p-1}\int_0^T\mathbb{E}\left(\sup_{0\le u\le s}|S_u^{j,M}-S_u^j|^2\right)ds\right.\\ &\quad +K_pT^{p-1}\int_0^T\mathbb{E}\left((S_s^{j,M}\eta_j(s,S_s^{j,M})-S_s^j\eta_j(s,S_s^j))^{2p}\right)ds\\ &\quad\left.+K_pT^{p-1}\beta_j^{2p}\int_0^T\mathbb{E}\left((S_s^{j,M}\sigma(s,I_s^M)-S_s^j\sigma(s,I_s))^{2p}\right)ds\right)\end{aligned}$$

Using assumption ($\mathcal{H}19$), one gets

$$\int_0^T\mathbb{E}\left((S_s^{j,M}\eta_j(s,S_s^{j,M})-S_s^j\eta_j(s,S_s^j))^{2p}\right)ds\le K_\eta^{2p}\int_0^T\mathbb{E}\left(\sup_{0\le u\le s}|S_u^{j,M}-S_u^j|^{2p}\right)ds.$$

Finally, by means of lemma 38 and assumptions ($\mathcal{H}17$) and ($\mathcal{H}18$),

$$\begin{aligned}
\int_0^T \mathbb{E}\left((S_s^{j,M}\sigma(s,I_s^M) - S_s^j\sigma(s,I_s))^{2p}\right) ds &\leq 2^{2p-1}\int_0^T \mathbb{E}\left((S_s^{j,M})^{2p}(\sigma(s,I_s^M)-\sigma(s,I_s))^{2p}\right) ds. \\
&\quad +2^{2p-1}\int_0^T \mathbb{E}\left((\sigma(s,I_s))^{2p}(S_s^{j,M}-S_s^j)^{2p}\right) ds \\
&\leq 2^{2p-1}C_{2p}^{\frac{1}{2}}K_{Lip}^{2p}T\sqrt{\mathbb{E}\left(\sup_{0\leq t\leq T}|I_t^M - I_t|^{4p}\right)} \\
&\quad +2^{2p-1}K_b^{2p}\int_0^T \mathbb{E}\left(\sup_{0\leq t\leq T}|S_s^{j,M}-S_s^j|^{2p}\right) ds
\end{aligned}$$

We deduce using Gronwall's lemma :

$$\mathbb{E}\left(\sup_{0\leq t\leq T}|S_t^{j,M}-S_t^j|^{2p}\right) \leq \widetilde{C}_T^j\sqrt{\mathbb{E}\left(\sup_{0\leq t\leq T}|I_t^M - I_t|^{4p}\right)}$$

where

$$\widetilde{C}_T^j = 6^{2p-1}K_pT^p\beta_j^{2p}C_{2p}^{\frac{1}{2}}K_{Lip}^{2p}\ e^{3^{2p-1}((r-\delta_j)^{2p}T^{2p-1}+K_pT^{p-1}K_\eta^{2p}+2^{2p-1}K_pT^{p-1}\beta_j^{2p}K_b^{2p})T}.$$

We conclude by Theorem 35 and the sublinearity of the square root function on $\mathbb{R}_+$.

We now turn to the L^{2p}-distance between I^M and $\overline{I}^M$:

$$\begin{aligned}
|I_t^M - \overline{I}_t^M|^{2p} &= \left|\sum_{j=1}^M w_jS_t^{j,M} - \sum_{j=1}^M w_jS_t^j\right|^{2p} \\
&\leq \left(\sum_{j=1}^M w_j|S_t^{j,M}-S_t^j|\right)^{2p} \\
&\leq \sum_{j_1=1}^M\cdots\sum_{j_{2p}=1}^M\prod_{k=1}^{2p} w_{j_k}|S_t^{j_k,M}-S_t^{j_k}|
\end{aligned}$$

So, using Hölder inequality, one has

$$\begin{aligned}
\mathbb{E}\left(\sup_{0\leq t\leq T}|I_t^M - \overline{I}_t^M|^{2p}|\right) &\leq \sum_{j_1=1}^M\cdots\sum_{j_{2p}=1}^M\left(\prod_{k=1}^{2p}w_{j_k}\right)\prod_{k=1}^{2p}\left(\mathbb{E}(\sup_{0\leq t\leq T}|S_t^{j_k,M}-S_t^{j_k}|^{2p})\right)^{\frac{1}{2p}} \\
&\leq \left(\sum_{j=1}^M w_j\right)^{2p}\max_{1\leq j\leq M}\widetilde{C}_T^j\left(\left(\sum_{j=1}^M w_j^2\right)^p + \right. \\
&\quad \left.\left(\sum_{j=1}^M w_j|\beta_j-\beta|\right)^{2p} + \left(\sum_{j=1}^M w_j|\delta_j-\delta|\right)^{2p}\right).
\end{aligned}$$

Chapitre 5

Estimation d'un modèle à intensités pour la gestion des risques. Extension aux modèles de *frailty* dynamique

Ce chapitre reprend un article écrit avec Jean-David Fermanian et Martin Delloye, dont une version abrégée a fait l'objet d'une publication dans la revue *Risk* (Delloye *et al.* [30]).

Abstract. We define a reduced-form credit portfolio model. Every rating transition is the outcome of several competing risks, which are independent conditionally on a set of observable explanatory variables and under the proportional hazards assumption. We perform a continuous time estimation using the Standard & Poor's historical database CreditPro. To allow stronger dependence levels between rating transitions for all the firms, we extend the model by adding a component of heterogeneity (a frailty), which is defined as an unobservable random process. Estimation issues are discussed and several empirical results are presented in both cases.

5.1 Introduction

Credit Portfolio models are key-tools for portfolio credit risk management, for economic capital calculations and for providing relevant inputs for Basel 2 regulatory requirements. Moreover, they are formally very closed to some pricing models for CDOs and basket derivatives.

Over the last few years, some models have emerged in the financial community and have been more or less successful. Among them, the most famous are Creditmetrics (JP Morgan, now Credit-Manager, marketed by RiskMetrics), Portfolio manager (KMV, now Moody's-KMV), CreditRisk+ (Credit Suisse Financial Products), CreditPortfolioView (McKinsey) and more recently Portfolio Risk Tracker (Standard & Poors). The reader is referred to the surveys of Crouhy *et al.* [26] or Koyluoglu and Hickman [72, 73]. It is well known that such models belong to the same class of factor models (see Schönbucher [110], Frey and McNeil [40]). The first three ones, the oldest ones, share particularly some similarities and can be mapped to each others (see Gordy [51]).

Intuitively, it is well understood that default probabilities depend on the overall economic situation. It has been established empirically for a long time : see Keenan *et al.* [66], Helwege and Kleiman [54], and more generally the annual reports from Standard and Poors or Moody's. Such an intuition has particularly led to the model CreditPortfolioView from McKinsey (cf. Wilson [132, 133]).

More generally, rating transitions (including the default state) are influenced by some macroeconomic variables although the internal rating process can introduce an additional noisy effect which is certainly not constant in time and cannot be taken into account easily.

Hence, several authors have tried to state and to estimate the rating changes process by finding some relevant explanatory variables. Most of them point out the influence of macroeconomic variables. For instance, Nickell *et al.* [94] estimate a multivariate probit model by working with individual Moody's rating histories. The authors underline the importance of the economic cycle, of the individual industry and of the geographic origin for explaining rating transitions. Kim [68] introduces the same type of model by considering a univariate synthetic economic "index" as a unique factor. We note that the latter index has been built for the transitions to default only, contrary to our approach. Lando [79] assumes one explanatory variable only (the short rate, which follow a Vasicek process) but without giving a true empirical justification. Otherwise, from Standard & Poor's rating histories, Bangia *et al.* [9] build two quarterly transition matrices, the first for economic "expansions", the other for "recessions". The authors conclude that the rating transition process can be considered as Markovian after conditioning by the state of the economy.

Our model is linked to the latter ones. It is a reduced-form model in the sense of Duffie and Singleton [36]. Our goal is to model simultaneously every rating transition in a consistent way. These transitions are assumed independent conditionally on some macroeconomic factors. These factors are explicit and easily observable. Such a model can be considered as a particular case of the Cox model (see Lando [79]). Nonetheless, much of the authors cited above are mainly concerned with the evaluation of credit sensitive security prices. That is why they calibrate their model with risky bond prices rather than individual rating histories. Contrary to them, our approach is fully historical.

To be specific, we seek to fit at best the past defaults and rating transitions events in order to predict conveniently future events under the physical probability. Nevertheless, we incorporate in our framework some current market or macroeconomic factors. Actually, we assume a type of

"stationarity" or "predictability", viz we assume that past events can explain risks in the future, which is the most usual risk measurement approach.

In order to generate flexible amounts of dependence between underlying rating transitions, we draw inspiration from the survival analysis literature to introduce common unobservable dynamic explanatory variables, called dynamic frailties.

Note that such a proportional hazards model can be valuable to price and/or hedge credit sensitive instruments, especially complex credit derivatives (CDO, basket, N-th to default, etc). In which case, it would be necessary to define other inputs, more market "informative" than ours (typically credit spreads or credit spread indices). Then, the model should be calibrated with some market prices like bonds or credit default swaps. Contrary to the commonly held view, such a "reduced-form" model is able to provide realistic correlation levels, as shown by Fermanian and Sbai [39] or Yu [136, 137].

The paper is organized as follows. In section 5.2, we describe the model and the successive steps for both estimation and forecast purposes and we provide some empirical results. In section 5.3, we explain how to estimate transition matrices for every time horizon and we provide some indicators to assess the quality of the fit. In section 5.4, we explain how to add more flexibility and more realism into the model by introducing a dynamic frailty process. We discuss estimation issues and we give some empirical results.

5.2 The basic model

We postulate a competing risks model. At every time t, any firm i is faced with $p-1$ underlying risks, which are the risks of rating changes (including default). In this study, we have worked with the usual aggregated rating scale: eight underlying credit states (AAA, AA, A, BBB, BB, B, CCC, with the S&P scale). Thus, p equals eight, but there is no theoretical hurdle for a larger number of categories. The associated time durations are assumed independent conditionally on some exogenous macroeconomic process and on some fixed and time-dependent idiosyncratic firm characteristics. All these variables are recorded in a firm-specific exogenous vector z. Formally, let $\alpha_{hji}(t)$ be the intensity of the transition from rating h to rating j for the firm i. As in Kavvathas [64] or Couderc and Renault [25], we set the traditional proportional hazards assumption. For every time t and every couple of transitions (h, j), $h \neq j$:

$$\alpha_{hji}(t|z) = \alpha_{hj0}(t) \exp(\beta'_{hj} z_{hji}(t)), \tag{5.1}$$

where α_{hj0} is an unknown deterministic function (the baseline hazard function), β_{hj} is an unknown vector of parameters and $z_{hji}(t)$ is the value of the i's covariate at time t. In this basic model case, the covariate is the same for every transition (h, j). This assumption can easily be relaxed.

Denote by $N_{hji}(t)$ the number of transitions from h towards j, for the firm i between the dates 0 and t. Thus, $dN_{hji}(t)$ is zero except when i makes a transition between ratings h and j at time t. Let $Y_{hi}(t)$ be an indicator variable valuing 1 when i is rated h just before the time t, and 0 otherwise.

The estimation of the unknown coefficients β_{hj} and α_{hj0} have been led by a full maximum likelihood procedure, as detailed in Andersen *et al.* [4] or Kavvathas [64]. With n firms, this

likelihood can be written as $\mathcal{L} = \prod_{i=1}^{n} \mathcal{L}_i$, where

$$\mathcal{L}_i = \left\{ \prod_t \prod_{j \neq h} (\alpha_{hji}(t|z))^{dN_{hji}(t)} \right\} \cdot \exp\left(- \sum_{(h,j)|j \neq h} \int_0^\infty Y_{hi}(u) \alpha_{hji}(u|z)\, du \right). \tag{5.2}$$

A right-censoring process disrupts our data set. This is due to the end of the observation period (31/12/2004 in our case) or caused by the entry into the category "Not Rated". The censoring variables have been assumed independent of the underlying risks. Even standard, this assumption could be questionable, especially for "Not Rated" outcomes. Nonetheless, there is no clear empirical link between the intensity towards the "Not Rated" category and some economic variables. Moreover, the defaults just after a "Not Rated" rating have been recorded in the database and have been taken into account in our likelihood. The left truncation can be neglected: by estimating the basic model on only the firms that enter in the sample after 1980, we get almost the same estimates as with all the firms.

The log-likelihood can be rewritten as

$$\ln \mathcal{L} = \sum_{j \neq h} \left\{ \sum_t \sum_{i=1}^{n} dN_{hji}(t)[\ln \alpha_{hj0}(t) + \beta'_{hj} z_{hji}(t)] - \sum_{i=1}^{n} \int_0^\infty Y_{hi}(u) \alpha_{hj0}(u) \exp(\beta'_{hj} z_{hji}(u))\, du \right\}.$$

Note that this likelihood can be split into a sum over every couple (h, j). Thus, we can estimate the parameters for any transition separately. This is very useful for practical implementations. Such likelihoods can be optimized easily with standard routines. In particular, when $\beta_{hj} = 0$ and the functions α_{hj0} are constant, the estimator of α_{hj0} can be obtained explicitly as

$$\widehat{\alpha}_{hj0} = \frac{\sum_{i=1}^{n} \sum_d N_{hji}(t)}{\sum_{i=1}^{n} \int_0^\infty Y_{hi}(u) du} = \frac{\text{Number of observed transitions from } h \text{ to } j}{\text{Occupation time in state } h}.$$

For convenience, we will assume that α_{hj0} are constant functions. This assumption can be relaxed, particularly with the methodology proposed by Couderc and Renault [25], but this would add complexity into the framework.

In terms of performances, the difference between the credit portfolio models are often due to calibration issues. To be more specific, the need for a historical database is crucial. Here we have worked with the CreditPro historical database from Standard and Poor's. This database contains individual firm rating histories recorded from 31/12/1980 to 31/12/2004: see some descriptive statistics in Tables 5.1 and 5.2. A large majority of firms are North-American. On the other side, a minority of firms are European or Asian. The latter category has been enriched significantly by S&P since the end of 90's, but now the rating history of European issuers is often a bit too short for a specific econometric study like ours. In the future, estimation procedures will be potentially feasible on some specific geographic areas or on some particular industries.

Our calibration procedure involves all the individual rating transitions. The exact date of every transition is inserted into the likelihood. This is an important difference with respect to the usual fits which rely on summaries like annual or even quarterly matrix transitions. Actually, such procedures lack a lot of information concerning quick rating transitions, especially just before default. Moreover, they cannot take into account a lot of infra-annual transitions. It will not be

our case, and the superiority of such "continuous time" fit is now widely recognized (see Lando and Skøderberg [80]).

For convenience and to avoid too much statistical noise, we have grouped several rating categories so that we get the usual eight grades:

- AAA $\mapsto$ AAA
- (AA, AA+, AA-) $\mapsto$ AA
- (A, A+, A-) $\mapsto$ A
- (BBB, BBB+, BBB-) $\mapsto$ BBB
- (BB, BB+, BB-) $\mapsto$ BB
- (B, B+, B-) $\mapsto$ B
- (CCC, CC, C +/-) $\mapsto$ CCC
- D $\mapsto$ D

Nonetheless, this is not a limitation of our modeling. An arbitrary number of grades is possible. In practice, we are restricted by the number of historically recorded rating transitions. Indeed, in order to get significant estimates for β_{hj}, one should observe at least ten transitions from rating h to rating j for every firm-specific modality.

On the other hand, in order to get statistically significant parameter estimates, it is valuable to group together rating transitions of the same type, under the assumption that the explanatory variables act similarly on all these transitions. Otherwise, β_{hj} can be set to zero when the number of transitions is too low (typically very outside the main diagonal, e.g. when $h = AAA$ and $j = CCC$).

Obviously, it is necessary to identify the main explanatory variables which drive the rating transitions and the defaults. Some of the previously cited authors have done a similar task in the past. Particularly, see an extensive analysis in Couderc and Renault [25], but concerning only the transitions towards default. For convenience, we have chosen a simple, perhaps crude, method: we focus on transition to default only, by assuming that the main relevant drivers for defaults will be relevant for the other rating transitions. Clearly, this assumption could be relaxed by leading a specific econometric analysis for every rating transition. This will certainly improve the empirical results and is left for further developments.

To be specific, using a usual linear regression, we have explained the monthly default rates for speculative-grade firms in the US between July 1986 and July 2002, as provided by the Federal Reserve Bank. We have retained four explanatory variables:

- the annual variation rate of the index of the industrial production in the US (denoted by "IPI" below),
- the annual variation rate of the Standard and Poor's 500 index (denoted by "SP"),
- the difference between the short term (3 months) and the long term (10 years) US government rates (denoted by "Slope")
- the 3-month short US government rate (denoted by "Short rate").

All these variables have been lagged. Indeed, the current values are not necessarily the most relevant ones to explain future default events. A change in the macroeconomic environment may take some time before affecting the observed default rates. These lags have been identified by a simple cross-correlation study. The "in sample fit" (by a simple ordinary least square regression) is excellent, with $R^2 = 87\%$: see Figure 5.1 page 151. Some descriptive statistics related to these variables are shown in Table 5.8.

The previous explanatory variables are "systemic". They aim to summarize the macroeconomic environment. Other firm-specific variables have been added as dummies:

– to be a firm which is based outside western Europe, US or Canada (denoted "not US-EU" below)
– to be a financial firm (denoted "Bank").

This may seem crude but other distinctions in terms of geographic area or industry were not found relevant. Also, there are many other powerful explanatory time-dependent firm-specific characteristics such as financial ratios, debt levels, equity or spread levels, etc. Unfortunately, they cannot be easily recorded and integrated in the informational system of a bank.

In addition, following some authors (Lando and Skøderberg [80], for instance), we take into account the firm-specific tendency towards upgrade or downgrade. Thus, at any time t, any firm will be in the state "up" or "down" if it has been upgraded or downgraded for the 12 months just before t. Therefore, we have attributed two time-dependent dummies to every firm, which try to capture some non-Markovian features of the rating process. We have recorded some descriptive statistics concerning the explanatory variables in Table 5.8 page 154.

The estimated coefficients β_{hj} are expanded below for every groups of transitions. Sometimes, some components of these coefficients do not appear because they are not significant (p-values larger than 1%), by usual univariate Wald tests. Nonetheless, we have tried to keep both interest rate covariates even when only one is significant. These tests are asymptotic and based on the empirical Hessian matrix. Then, we have estimated the full model again by retaining only the significant variables. This is a standard empirical procedure for maximum likelihood estimation.

For instance, we get that the monthly intensity at time t for a transition from rating BBB to BB is

$$\alpha_{BBB\to BB}(t) = \exp\left(-5.3422 - 2.838 * \mathrm{IPI}_t - 1.177 * \mathrm{S\&P}_t - 0.0043 * \mathrm{slope}_t + 0.0243 * \text{short rate}_t \right.$$
$$\left. + 0.0243 * \text{short rate}_t + 0.128 * \mathbf{1}(\text{neither US, Canada, west. Europe}) - 0.574 * \mathbf{1}(\text{upgraded last year})\right)$$

The latter equation can be read as follows: in the first order, when the explanatory variable "variation rate of IPI" goes up 1% from the previous month, the transition intensity from BBB to BB goes down around 2.8%. The slope of the interest rates and the short rate are recorded as percentages. Therefore, when the short rate is becoming 4% after 3% previously, these transition intensity goes up 2.4% approximately.

– Transitions from CCC to D.

parameter	**estimation**	**st. deviation**	**est./st.d.**	**p. value**
α_0(CCC $\rightarrow$ D)	-3.1731	0.1374	-23.100	0.0000
SP	-1.0696	0.1833	-5.836	0.0000
Slope	-0.1853	0.0347	-5.347	0.0000
Short rate	-0.0843	0.0187	-4.519	0.0000
Down	1.3134	0.0667	19.690	0.0000

– Transitions from B to D.

parameter	**estimation**	**st. deviation**	**est./st.d.**	**p. value**
α_0(B $\rightarrow$ D)	-6.0303	0.0661	-91.256	0.0000
IPI	-11.6772	1.4569	-8.015	0.0000
Down	1.0823	0.1337	8.097	0.0000

– Transitions from BB to B and from B to CCC.

parameter	**estimation**	**st. deviation**	**est./st.d.**	**p. value**
α_0(BB $\rightarrow$ B)	-4.2195	0.0796	-53.022	0.0000
α_0(B $\rightarrow$ CCC)	-4.2818	0.0793	-54.005	0.0000
IPI	-7.2104	0.5858	-12.309	0.0000
Slope	-0.1563	0.0200	-7.802	0.0000
Short rate	-0.0622	0.0098	-6.351	0.0000
not US-EU	0.1102	0.0601	1.833	0.0668
Down	1.1474	0.0499	22.984	0.0000
Up	-0.8694	0.1908	-4.556	0.0000

– Transitions from AA to A, from A to BBBB, and from BBB to BB.

parameter	**estimation**	**st. deviation**	**est./st.d.**	**p. value**
α_0(AA $\rightarrow$ A)	-4.9084	0.0768	-63.929	0.0000
α_0(A $\rightarrow$ BBB)	-5.2546	0.0726	-72.390	0.0000
α_0(BBB $\rightarrow$ BB)	-5.3422	0.0716	-74.603	0.0000
IPI	-2.8389	0.5489	-5.172	0.0000
SP	-1.1774	0.1178	-9.996	0.0000
Slope	-0.0043	0.0176	-0.242	0.8087
Short rate	0.0243	0.0084	2.894	0.0038
not US-EU	0.1285	0.0578	2.223	0.0262
Up	-0.5749	0.1273	-4.517	0.0000

– One notch upgrades

parameter	**estimation**	**st. deviation**	**est./st.d.**	**p. value**
α_0(A → AA)	-7.1827	0.0921	-77.999	0.0000
α_0(BBB → A)	-6.3918	0.0824	-77.544	0.0000
α_0(BB → BBB)	-6.0278	0.0828	-72.837	0.0000
α_0(B → BB)	-6.0246	0.0819	-73.547	0.0000
IPI	3.6968	0.5936	6.227	0.0000
Slope	0.1131	0.0187	6.038	0.0000
Short rate	0.0816	0.0086	9.538	0.0000
not US-EU	-0.1278	0.0711	-1.798	0.0722
Bank	0.4349	0.0552	7.879	0.0000
Down	-0.5187	0.0992	-5.228	0.0000
Up	-0.4994	0.1369	-3.649	0.0003

– Two notches downgrades

parameter	**estimation**	**st. deviation**	**est./st.d.**	**p. value**
α_0(AA → BBB)	-9.2168	0.2627	-35.089	0.0000
α_0(A → BB)	-9.3431	0.2378	-39.286	0.0000
α_0(BBB → B)	-8.6270	0.2133	-40.452	0.0000
α_0(BB → CCC)	-8.2019	0.2111	-38.857	0.0000
IPI	3.5299	1.4842	2.378	0.0174
SP	-1.0205	0.3450	-2.958	0.0031
Slope	0.1947	0.0500	3.898	0.0001
Short rate	0.1585	0.0225	7.059	0.0000
not US-EU	0.1197	0.1956	0.612	0.5405
Bank	-0.9690	0.2115	-4.582	0.0000
Down	0.6232	0.1758	3.545	0.0004

– Two notches upgrades

parameter	**estimation**	**st. deviation**	**est./st.d.**	**p. value**
α_0(A → AAA)	-11.8202	0.4182	-28.263	0.0000
α_0(BBB → AA)	-10.2174	0.3298	-30.977	0.0000
α_0(BB → A)	-9.3849	0.3206	-29.271	0.0000
α_0(B → BBB)	-9.5039	0.3213	-29.575	0.0000
α_0(CCC → BB)	-8.6390	0.3644	-23.705	0.0000
Slope	0.2096	0.0701	2.992	0.0028
Short rate	0.1637	0.0315	5.193	0.0000
not US-EU	-1.8942	0.5182	-3.655	0.0003
Bank	1.5888	0.1904	8.346	0.0000

– Transitions from AAA to AA

parameter	estimation	st. deviation	est./st.d.	p. value
α_0(AAA → AA)	-5.9351	0.2339	-25.377	0.0000
IP	4.5874	1.6512	2.778	0.0055
Slope	0.0339	0.0514	0.659	0.5096
Short rate	0.1003	0.0240	4.186	0.0000
not US-EU	0.4953	0.1790	2.767	0.0057
Up	-1.3025	0.7096	-1.836	0.0664

– Transitions from AA to AAA

parameter	estimation	st. deviation	est./st.d.	p. value
α_0(AA → AAA)	-7.6137	0.1959	-38.868	0.0000
IP	6.9505	3.4019	2.043	0.0410
SP	-0.9643	0.7613	-1.267	0.2053
not US-EU	0.7048	0.3080	2.289	0.0221
Bank	-0.5249	0.2992	-1.755	0.0793
Down	0.7948	0.5163	1.539	0.1237

– Transitions from CCC to B

parameter	estimation	st. deviation	est./st.d.	p. value
α_0(CCC → B)	-4.9530	0.3049	-16.244	0.0000
SP	-1.1324	0.4460	-2.539	0.0111
Slope	0.3485	0.0804	4.335	0.0000
Short rate	-0.0583	0.0420	-1.387	0.1653

To our knowledge, this paper is the first one that provides such estimates in a full competing risk framework. Briefly, we analyze these results qualitatively:

– For almost every transitions, when the IPI or the S&P indexes increase, the probability of downgrades (including default) goes down. This is particularly the case for the transition from B to D. Conversely, a direct transition from B to default is an exceptional event which is less sensitive with respect to the macroeconomic situation.
– The explanatory power of our variables is stronger for downgrades than for upgrades, even if an increase of the IPI rate has a significant impact on the latter.
– The transitions from AAA towards AA and conversely seem to be unconnected with the other rating transitions. For example, the IPI rate seems to "boost" the corresponding intensities for both transitions, contrary to the S&P index.
– It is relatively difficult to analyze the effect of the two "interest rates related" variables. Globally, when the rates go up, the macroeconomic environment is better. Thus, speculative-grade firms are downgraded less frequently. Surprisingly, the opposite effect occurs with the investment-grade firms. This may show their tendency to take more risk in such an environment. Moreover, the two-notch moves are more frequent when the rates are high, as if the temperature of the economy was increasing: more liquidity constraints for the weakest firms, and more opportunities for the strongest ones.
– Banks are more easily upgraded than the other firms and can downgrade just as easily as for any other sector.
– The rating process seems to be harder for the firms located outside the US and western Europe. Such firms are more often downgraded and less often upgraded, although their default rates are not higher when they are rated CCC.
– When a firm has been downgraded for less than a year, its downgrade (including default)

probability increases significantly. Such an effect is less sensitive when the firm has been upgraded for less than a year. Surprisingly, when a firm has been upgraded by one notch, its current probability for additional upgrades falls. In other words, it is easier to be downgraded quickly than upgraded quickly. This phenomenon can be explained by the way the rating agencies work (annual revisions for "without-any-problems" firms).

5.3 Computation of the transition matrices and Tests in sample

After having estimated the matrices of transition intensities for firm i by $\widehat{I}_i(t) = [\widehat{\alpha}_{hji}(t)]_{1\leq h,j\leq p}$, we deduce an estimate for monthly transition matrices at time t by $\widehat{P}_i(t, t+1) = Id_p + \widehat{I}_i(t)$, where Id_p is the (p, p)-identity matrix. Obviously, we have set $\widehat{\alpha}_{hhi} = -\sum_{j\neq h} \widehat{\alpha}_{hji}(t)$.

It is now possible to calculate transition matrices between two arbitrary dates t_1 and t_2 (in months) by simply multiplying successive monthly transition matrices:

$$\widehat{P}_i(t_1, t_2) = \prod_{k=t_1}^{t_2-1} \widehat{P}_i(k, k+1)$$

This procedure is justified in the usual counting processes theory (see Andersen et al., 1993).

To assess the quality of the model, we calculate the mean annual transition matrix 1981-2004 as computed by our model, using the past values of the relevant macroeconomic explanatory variables. We compare this matrix with the S&P annual transition matrix as provided by CreditPro: see Tables 5.4 and 5.3. The fit is good, especially concerning the main diagonal and the "one notch" downgrades or upgrades. Nonetheless, the yearly default rates are too low, especially for the categories BBB and BB. This is due to a strong non Markovian feature of these particular rating transitions. In other words, it has been observed in the past a "too high" number of such transitions with respect to the model. To stay conservative, it is always possible to introduce a floor level for such default rates, for instance 0.03% per year as proposed by Basel 2 (cf. [99], paragraph 285). It is also possible to get round this problem by increasing the time unit: instead of monthly intensities, consider quarterly or even six-monthly intensities. Note that, due to the continuous time estimation, all the transition probabilities in Table 5.4 are non zero (even if they are very low, sometimes).

Moreover, we have focused on the monthly default rates themselves. They are compared with those obtained empirically from the S&P database. We obtain Figure 5.2. After smoothing the empirically observed default rates, the fit is very good. Thus, our model is able to recover the fluctuations of the historical default rates which are due to the economic cycle.

5.4 Extension to frailty models

In the statistical literature, it is well recognized that some unobservable random variables can affect the underlying endogenous process. Indeed, we could have forgotten some relevant macroeconomic factors, or some firm-specific characteristics in the explanatory variables vector z_{hji} (see equation (5.1)). For instance, some authors include the inflation rate or the unemployment rate in z_{hji}. We may also imagine that variables such as the skills of the firm management, or the expenses in R&D could influence the future firm performances. To deal with this issue, we extend the basic

model by introducing a so-called "frailty" variable γ_{hjit}. By nature, such variables cannot be observed. They summarize the effects of all the (systemic or specific) variables that were forgotten into the basic model.

Frailty models have been used for a long time in survival analysis (see Clayton and Cuzick [23] or Hougaard [56] for a survey), and have been applied in many fields: biostatistics, microeconomics, reliability, insurance,etc. In the finance area, they have been recently proposed for credit risk modelling: Metayer [91] introduces frailties as static multiplicative factors in a Cox model, allowing the calculation of closed-form likelihood criteria and Schönbucher [111] assumes a strong contagion effect between firms whose (static) frailties are strongly correlated. In Fermanian and Sbai [39], it is shown that dealing with frailties in intensity-based credit risk models induces the same dependence levels between default events as in the usual Merton-style models. Such type of conclusions appeared also in Schönbucher [111].

To be more specific, we will differ from the current literature by considering a frailty process rather than "static" frailties : for every firm i, every time t and every couple of transitions (h, j), $h \neq j$, we now assume

$$\alpha_{hji}(t|z) = \gamma_{hjit}\alpha_{hj0}\exp(\beta'_{hj}z_{hji}(t)). \tag{5.3}$$

Thus, the "conditional" likelihood (the likelihood of the transitions knowing the frailty values) is the product of n "firm specific conditional" likelihoods: $\mathcal{L}^c = \prod_{i=1}^n \mathcal{L}_i^c$ where

$$\mathcal{L}_i^c = \left\{ \prod_t \prod_{j\neq h} \left(\gamma_{hjit} e^{\alpha_{hj0}+\beta'_{hj}z_{hji}(t)}\right)^{dN_{hji}(t)} \right\} e^{-\sum_{j\neq h}\int_0^\infty Y_{hi}(u)\gamma_{hjiu}e^{\alpha_{hj0}+\beta'_{hj}z_{hji}(u)}du}. \tag{5.4}$$

The full likelihood is the expectation of this conditional likelihood with respect to the law of the frailties. We assume the frailties do not depend on i. To lighten the notations, we consider a single group of transitions (for example, only the one notch downgrades). Therefore, the sub-indixes h, j can be removed. That is why the frailty process is denoted now simply γ_t. We get the full likelihood:

$$\begin{aligned} \mathcal{L} &= \mathbb{E}_\gamma(\mathcal{L}^c) \\ &= C(\beta)\,\mathbb{E}_\gamma\left(\prod_{t=1}^{T_0}\gamma_t^{\sum_{i=1}^n\sum_{j\neq h}\Delta N_{hji}(t)}\, e^{-\gamma_t\sum_{i=1}^n\sum_{j\neq h}\int_{t-1}^t Y_{hi}(u)e^{\alpha_{hj0}+\beta'_{hj}z_{hji}(u)}du}\right), \end{aligned} \tag{5.5}$$

where

$$C(\beta) = \prod_{i=1}^n\prod_t\prod_{j\neq h} e^{\left(\alpha_{hj0}+\beta'_{hj}z_{hji}(t)\right)dN_{hji}(t)}.$$

Here, we have chosen an annual time unit. It means that the frailties are constant during a whole year. The time T_0 denotes the historical length of our database, that is, $T_0 = 23$. Because of the dynamic feature of the frailties, the previous equation (5.5) cannot be simplified: no closed-form formulas exist, except in the special case of constant frailties.

Dynamic frailties are intuitively natural: like the observed macro factors z_{hji}, the unobserved factors that drive the credit risk should be time-dependent. This is not a detail. Indeed, in our multi-period framework, it is the only way to increase significantly the dependence levels between

rating changes with respect to the basic model. Actually, by drawing independently different frailties from one period of time to another, a "law of large numbers" implies some type of compensation between successive periods: what we get in hectic conjuncture (in terms of losses) can be lost hereafter during quiet periods. Therefore, it is important to introduce some inertia in the frailty process $(\gamma_t)_{t=1,\ldots,T}$. We choose the following specification :

$$\gamma_1 = \tilde{\gamma}_1, \quad \gamma_t = \gamma_{t-1}\tilde{\gamma}_t,$$

for every time $t = 2, \ldots, T$.

The random variables $\tilde{\gamma}_t$ are drawn independently for every t. They follow a gamma law whose single parameter is denoted by α: we impose that

$$\mathbb{E}[\tilde{\gamma}_t] = 1 \text{ and } Var(\tilde{\gamma}_t) = 1/\alpha.$$

Thus, the frailties add some additional variability on the hazard rates, by preserving their means (conditionally on the macroeconomic covariates). Such dynamic frailty models are seldom found in the literature. Nonetheless, Yue and Chan [138] have studied the same type of autoregressive process as ours. See also Paik *et al.* [100] or Yau and McGilchrist [135] for other types of frailty processes. Notably, Koopman *et al.* [71] have proposed the same family of models as ours independently. However, we differ concerning the estimation procedure (based on the Kalman filter in their case and on the Monte Carlo Markov Chain (MCMC) here).

In such models, the main difficulty lies in inference and the lack of tractable formulas. Particularly, no more closed-form formulas are available for the maximum likelihood criterion, contrary to constant (static) gamma frailty components.

Denote by p the density of the vector of frailties $\gamma^{T_0} = (\gamma_1, \ldots, \gamma_{T_0})'$. Clearly,

$$p(d\gamma^{T_0}) = \prod_{t=1}^{T_0} g\left(\alpha, \frac{\alpha}{\gamma_{t-1}}\right)(\gamma_t) \; d\gamma_1 \cdots d\gamma_{T_0},$$

where $g(\alpha, \beta)$ denotes the density of a gamma random variable:

$$g(\alpha, \beta)(x) = \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x} \mathbf{1}_{\mathbb{R}^+}(x).$$

We have set $\gamma_0 = 1$ as a convention. Coming back to the likelihood (5.5), we can go one step further: $\mathcal{L}(\theta) = \displaystyle\int_{\mathbb{R}^{T_0}} \overline{\mathcal{L}}(\theta)\, p(d\gamma^{T_0})$, where

$$\overline{\mathcal{L}}(\theta) = C(\beta) \prod_{t=1}^{T_0} \gamma_t^{\sum_{i=1}^n \sum_{j\neq h} \Delta N_{hji}(t)} \, e^{-\gamma_t \sum_{i=1}^n \sum_{j\neq h} \int_{t-1}^t Y_{hi}(u) e^{\alpha_{hj0} + \beta'_{hj} z_{hji}(u)} du}.$$

$\overline{\mathcal{L}}(\cdot)$ is the so-called "complete" likelihood (the likelihood associated with all the rating transitions and frailties, if the frailties were observable). It is a function of θ, the vector of all the model parameters. The vector θ contain the unknown α but also the coefficients α_{hj0} and β_{hj}.

The estimation of the model implies the full maximization of $\mathcal{L}$ with respect to θ. This is clearly challenging. To tackle this issue, we propose an 'expectation-maximization' inference method based

on Monte Carlo Markov Chains simulation techniques. The EM algorithm is a popular and efficient approach to maximum likelihood estimation for incomplete data (see the survey of McLachlan and Krishnan [88]). It induces an iterative optimization method which generates a sequence of estimated parameters (θ_k). Starting from an "arbitrary" θ_0, and assuming we have found $\theta_1, \ldots, \theta_k$, we need to maximize the following $Q(\cdot|\theta_k)$ criterion defined as

$$Q(\theta|\theta_k) = \int_{\mathbb{R}^{T_0}} \ln(\overline{\mathcal{L}}(\theta)) p(d\gamma^{T_0}|Y, \theta_k), \tag{5.6}$$

where $p(\cdot|Y, \theta_k)$ denotes the density of the frailties vector γ^{T_0} knowing all the observations Y (all the rating transitions) and assuming the value of our parameter is θ_k. At each step, we approximate the integral in equation (5.6) by a sum, by a usual Monte Carlo procedure. This is the usual Simulated EM approach, called SEM (Celeux and Diebolt [21]). Therefore, we need to draw in the conditional law $p(\cdot|Y, \theta_k)$.

This last step can be done by MCMC techniques (see Robert and Casella [104] for example). Here, a Hastings-Metropolis algorithm with random walk is proposed: knowing $\gamma_s^{T_0}$,

1. Generate $y_s \sim q(\cdot - \gamma_s^{T_0})$.
2. Generate

$$\gamma_{s+1}^{T_0} = \begin{cases} y_s & \text{with probability } \rho = \min\left(1, p(y_s|Y, \theta_k)/p(\gamma_s^{T_0}|Y, \theta_k)\right), \\ \gamma_s^{T_0} & \text{with probability } 1 - \rho. \end{cases}$$

The instrumental distribution q has been chosen to be uniform on $[-0.1, 0.1]^{T_0}$. Note that, in our case, the MCMC probability ρ is relatively easy to compute since, for every parameter $\overline{\theta}$, we have

$$\frac{p(y_s|Y, \overline{\theta})}{p(\gamma_s^{T_0}|Y, \overline{\theta})} = \prod_{t=1}^{T_0} \left(\frac{y_s(t)}{\gamma_s^{T_0}(t)}\right)^{\nu(t)+\overline{\alpha}-1} \left(\frac{\gamma_s^{T_0}(t-1)}{y_s(t-1)}\right)^{\overline{\alpha}} e^{b(t)\left(\gamma_s^{T_0}(t) - y_s(t)\right) - \overline{\alpha}\left(\frac{y_s(t)}{y_s(t-1)} - \frac{\gamma_s^{T_0}(t)}{\gamma_s^{T_0}(t-1)}\right)},$$

where

$$\begin{cases} \nu(t) &= \sum_{i=1}^n \sum_{j \neq h} \Delta N_{hji}(t) \\ b(t) &= \sum_{i=1}^n \sum_{j \neq h} \int_{t-1}^t Y_{hi}(u) e^{\overline{\alpha}_{hj0} + \overline{\beta}'_{hj} z_{hji}(u)} du. \end{cases}$$

By repeating the same procedure S times, we simulate a Markov chain $(\gamma_s^{T_0})_{s=1,\ldots,S}$ whose stationary law is $p(\cdot|Y, \theta_k)$. Hence, we approximate the criterion $Q(\theta|\theta_k)$ by

$$\tilde{Q}(\theta|\theta_k) = \frac{1}{S} \sum_{s=1}^{S} \ln(\overline{\mathcal{L}}(\theta))(\gamma_s^{T_0}),$$

which we maximize with respect to the parameter θ to get θ_{k+1}. This idea (MCMC random draws inside a SEM procedure) has already been proposed in the statistical literature: see Diebolt and Ip [32], for instance.

This frailty model clearly extends the basic model of section 5.2. To simplify the analysis and to avoid a type of dilution of the frailty effects among the different group of rating transitions, we have reduced their number. We have kept only two groups: one notch downgrades and one notch upgrades. The estimated coefficients β_{hj} are qualitatively close to those obtained in section 5.2, so they are not recorded again here. Our estimated α is 23.6 in the case of downgrades, and 52.0

for the upgrades. Such values seem to be large and generate small $\tilde{\gamma}_t$ variances. Nonetheless, in the long run, the process γ_t can reach relatively high values. For instance, when α equals 100, the standard deviation of γ_t, for $t = 10, 20$ and 30 is respectively 0.33, 0.46 and 0.58. For $t = 10$ and larger, the γ_t quantile at 95% is larger than 1.61 (See Table 5.7).

Hence, by introducing dynamic frailties, the dependence between rating transitions and default events can be significantly increased, especially in the long run. This is particularly relevant for multi-period economic capital calculations. As an illustration, we have considered a simple portfolio (50 firms, different ratings from AAA to CCC), with long term (30 years) constant equal exposures. We have simulated 1,000 random paths of the observable explanatory random variable (assuming they follow a first-order Vectorial Auto Regressive process $VAR(1)$). For every previous random path, 10 frailty paths are drawn. Then, we calculate the losses due to default events between today and the time horizon, along the 10,000 drawn scenarios. We obtain Figure 5.3. Clearly, the tails are potentially a lot fatter when dynamic frailties are introduced. For the basic model, the shape of the loss distribution appears Gaussian (even if this is not really the case), which is surprising. Actually, we have noticed that these shapes depend strongly on the z_t process we choose in practice : mean reversion or not, the degree p in $VAR(p)$, etc. So, by choosing other econometric processes, the tails can become fatter, even for the basic model.

For the sake of comparison, we have calculated the mean default rates at several horizons as provided by our basic model, the frailty model and by S&P (historical rates): see Table 5.5. Globally, our models tend to underestimate the default rates for investment grade firms. This is probably due to some non Markovian features of the historical rating processes. A simple way to tackle this issue would be to increase the time unit: instead of estimating monthly transition matrices, consider quarterly or even semi-annual matrices. But we would then lose the nice features of the continuous-time description.

Finally, we have compared correlations between the default events: see Table 5.6 page 152. As shown in De Servigny and Renault [28], the correlation estimates are fragile. Thus, we have followed the same methodology as S&P. By simulating 10,000 rating transitions over one, two and five years for every cohort, we find a coherence with the S&P correlation levels, particularly with the frailty model and for short term horizons. The historical correlation levels are relatively well recovered, which was *a priori* a difficult task.

It is important to note that the extra-variability induced by the frailty processes allows us to get a significant range of default rates and correlation levels. Thus, for a significant number of trajectories, these quantities are equal or even larger than those observed historically. This is important in a Credit VaR or economic capital perspective. The frailty model will be more conservative than the basic one.

Finally, note that, for all these empirical comparisons, we have controlled the sample effects: for every cohort, we have considered the same number of firms as in the S&P database in terms of ratings and other idiosyncratic characteristics (bank or not, US-EU or not, up or down). As S&P does, we calculate our statistical indicators by some weighted means with respect to the sizes of these cohorts.

5.5 Conclusion

We have explained how to specify and estimate a reduced-form credit portfolio model in a full competing risks approach. Extensions towards dynamic frailty models allow us to take into account unobservable explanatory variables in a nice way and generate potentially a large range of dependence levels between rating changes. The "in sample" fit with the S&P historical database is satisfying, especially concerning the correlation levels between default events. For economic capital calculations, this model can be applied by assuming the explanatory random variables follow a particular random process, for instance vectorial auto regressive. We have observed that the performances of the model can be relatively sensitive for different choices of the latter process. But, even if such point is important in practice, it does not jeopardize the relevance of our approach. Moreover, our model can be used straightforwardly as a predictor of the credit market, under some stress-tests assumptions or some macroeconomic forecasts.

Industry	US, Canada	West. Eur.	Latin Amer.	Japan, Austr.	Others	Total
Automotive, metal	881	126	43	41	42	1131
Consumer, service sector	1081	116	58	50	20	1325
Energy, natural resources	454	41	20	15	31	561
Financial Institutions	982	495	104	145	298	2024
Forest, homebuilders	295	43	27	11	13	389
Health care, chemicals	486	69	13	7	13	588
High tech., computers	335	23	2	14	23	397
Insurance	504	100	0	33	46	683
Leisure time, media	692	62	28	16	12	810
Real estate	239	28	6	25	12	310
Telecommunications	327	60	32	8	59	486
Transportation	300	64	11	32	33	440
Utility	608	138	57	64	51	918
Total	7184	1365	401	461	651	10062

Table 5.1: Distribution of firms in the Standard and Poor's database Credit Pro (july 2003)

in %	AAA	AA	A	BBB	BB	B	CCC	D	Censored and NR	Total
AAA	0	280	14	3	2	0	0	0	251	550
AA	71	0	916	40	6	5	1	0	826	1865
A	12	419	0	1212	64	35	2	0	2132	3876
BBB	5	35	686	0	933	111	7	8	2398	4183
BB	4	10	47	669	0	1197	124	28	1831	3910
B	0	7	22	40	701	0	1207	321	2035	4333
CCC	1	0	6	8	17	194	0	1058	262	1546
Total	93	751	1691	1972	1723	1542	1341	1415	9735	20263

Table 5.2: Number of transitions 1/1/1981-31/12/2004 as observed in the historical S&P database CreditPro 7.0

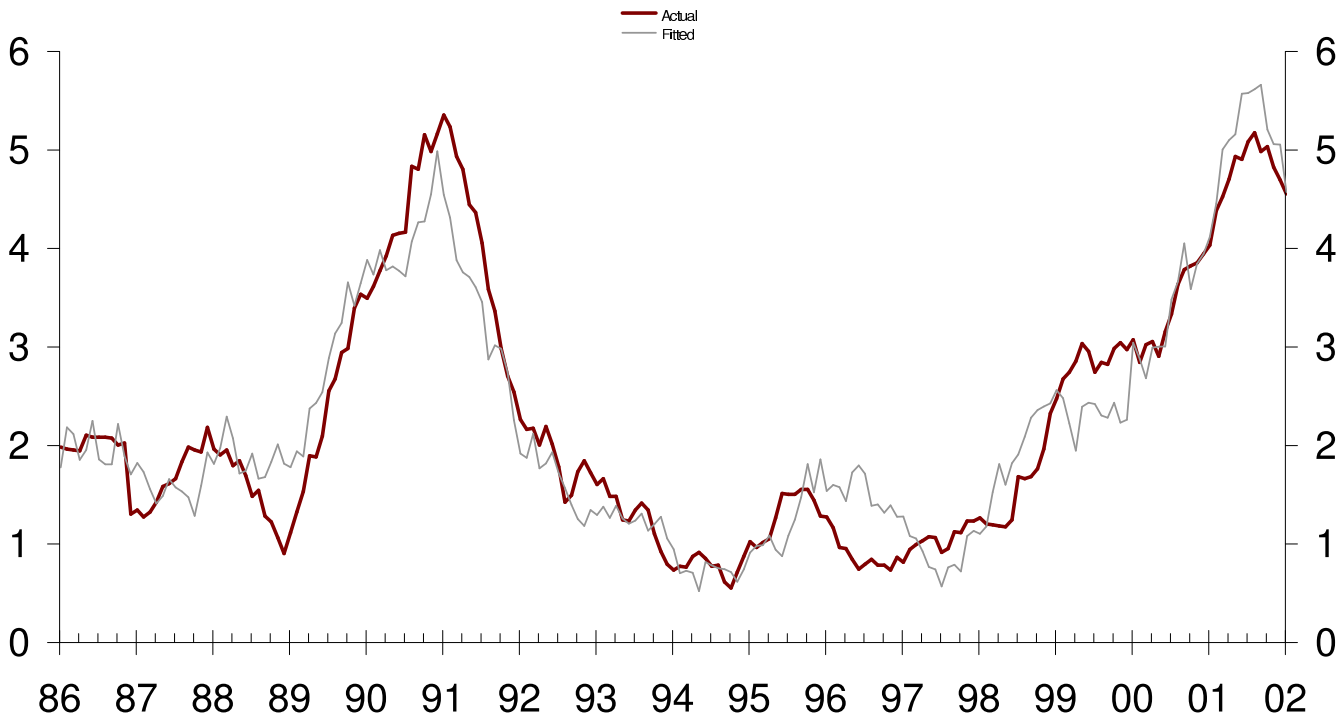

Figure 5.1: Monthly default rates (US firms)

in %	**AAA**	**AA**	**A**	**BBB**	**BB**	**B**	**CCC**	**D**
AAA	91.76	7.39	0.68	0.11	0.06	0.00	0.00	0.00
AA	0.61	90.76	7.91	0.58	0.08	0.05	0.01	0.02
A	0.06	1.97	91.70	5.61	0.45	0.19	0.02	0.07
BBB	0.03	0.23	3.85	89.88	4.99	0.85	0.09	0.11
BB	0.03	0.09	0.48	5.13	83.81	8.58	1.29	0.69
B	0.00	0.06	0.19	0.46	5.01	83.07	7.28	4.58
CCC	0.04	0.01	0.22	0.31	0.70	6.28	57.81	36.53

Table 5.3: Yearly transition matrix 1981-2004 as provided by our model

in %	**AAA**	**AA**	**A**	**BBB**	**BB**	**B**	**CCC**	**D**
AAA	91.83	7.50	0.48	0.12	0.06	0.00	0.00	0.00
AA	0.65	90.24	8.30	0.62	0.05	0.12	0.02	0.01
A	0.05	2.20	91.05	5.98	0.46	0.18	0.04	0.05
BBB	0.03	0.24	4.26	89.04	5.01	0.87	0.22	0.33
BB	0.03	0.09	0.39	5.91	82.84	8.26	1.12	1.36
B	0.00	0.08	0.24	0.33	5.67	82.08	4.97	6.63
CCC	0.10	0.00	0.30	0.50	1.59	10.43	53.03	34.06

Table 5.4: Yearly transition matrix 1981-2004 as provided by S&P (source: CreditPro 7.0)

in %	1 Year			2 Years			5 Years		
	Basic	Frailty	S&P	Basic	Frailty	S&P	Basic	Frailty	S&P
AAA	0.000	0.000 (10^{-4})	0.00	0.001	0.001 (0.001)	0.000	0.016	0.017 (0.024)	0.098
AA	0.002	0.002 (0.001)	0.01	0.010	0.011 (0.012)	0.041	0.080	0.104 (0.267)	0.319
A	0.007	0.007 (0.006)	0.04	0.037	0.043 (0.066)	0.135	0.279	0.335 (0.681)	0.654
BBB	0.107	0.112 (0.057)	0.33	0.335	0.379 (0.390)	0.903	1.641	1.704 (1.917)	3.266
BB	0.686	0.706 (0.560)	1.27	2.212	2.174 (1.918)	3.824	7.806	7.052 (4.978)	12.10
B	4.588	4.503 (2.321)	6.06	10.46	9.720 (4.984)	13.22	24.08	21.12 (8.999)	26.40
CCC	36.53	28.49 (12.90)	30.5	48.11	41.91 (14.62)	39.50	67.44	59.17 (14.13)	52.68

Table 5.5: Default rates provided by the "basic" model, the frailty model (1000 trials, standard deviations into brackets) and S&P statistics (CreditPro 7.0).

in %	1 year			2 years			5 years		
	IG	SG	crossed	IG	SG	crossed	IG	SG	crossed
S&P	0.10	1.37	0.26	0.28	2.13	0.56	0.31	2.68	0.67
Basic	0.003	1.02	0.12	0.025	1.628	0.447	0.010	0.183	-0.147
Frailty									
Mean	0.0104	1.208	0.175	0.089	2.130	0.708	0.174	1.340	0.472
$q5\%$	0.0012	0.388	-0.011	0.005	0.563	-0.033	0.006	0.078	-0.361
$q25\%$	0.0015	0.572	0.024	0.007	0.938	0.0440	0.021	0.340	-0.266
$q50\%$	0.0021	0.798	0.062	0.0115	1.343	0.165	0.037	0.805	-0.203
$q75\%$	0.0047	1.423	0.157	0.0357	2.415	0.577	0.071	1.381	-0.008
$q95\%$	0.0296	2.533	0.562	0.261	5.001	2.475	0.271	3.441	1.460

Table 5.6: Correlations between default events at several horizons as provided by our models and by S&P. In the frailty case, we get a sample of correlations (one for each frailty random path). We have detailed the means and the quantiles of these samples.

	Mean	St. dev.	Quantile 95%	Maximum
1 year	0.999	0.101	1.171	1.431
2 years	0.999	0.142	1.248	1.694
3 years	1.000	0.175	1.310	1.987
4 years	1.003	0.203	1.370	1.958
5 years	1.003	0.229	1.415	2.318
6 years	1.004	0.252	1.460	2.345
7 years	1.008	0.274	1.509	2.459
8 years	1.009	0.294	1.548	2.825
9 years	1.009	0.314	1.579	3.297
10 years	1.008	0.332	1.615	3.225
11 years	1.008	0.346	1.641	3.419
12 years	1.007	0.360	1.672	3.582
13 years	1.006	0.373	1.699	3.996
14 years	1.004	0.388	1.730	4.630
15 years	1.003	0.402	1.768	4.962
16 years	1.003	0.414	1.781	4.272
17 years	1.002	0.427	1.804	4.455
18 years	1.002	0.438	1.824	5.206
19 years	1.002	0.452	1.841	5.542
20 years	1.001	0.464	1.873	5.670
21 years	1.000	0.479	1.894	5.314
22 years	1.001	0.491	1.916	4.988
23 years	1.002	0.503	1.944	5.582
24 years	1.003	0.516	1.973	5.831
25 years	1.001	0.526	2.008	6.153
26 years	1.001	0.539	2.032	6.479
27 years	1.001	0.551	2.059	6.052
28 years	1.001	0.562	2.083	6.408
29 years	1.001	0.569	2.095	6.729
30 years	1.003	0.581	2.121	7.572

Table 5.7: Statistical indicators of the dynamic frailty process: simulation of 1,000 paths with $\alpha = 100$.

	IPI	SP	Slope	Short rate
Mean	0.0262	0.129	0.499	8.10
Variance				
IPI	0.00150	0.00086	0.0297	-0.0523
S&P	0.00086	0.0258	0.0359	0.0479
slope	0.02970	0.0359	3.3418	-5.303
short rates	-0.05235	0.0479	-5.3030	14.349

Table 5.8: Description of the explanatory variables

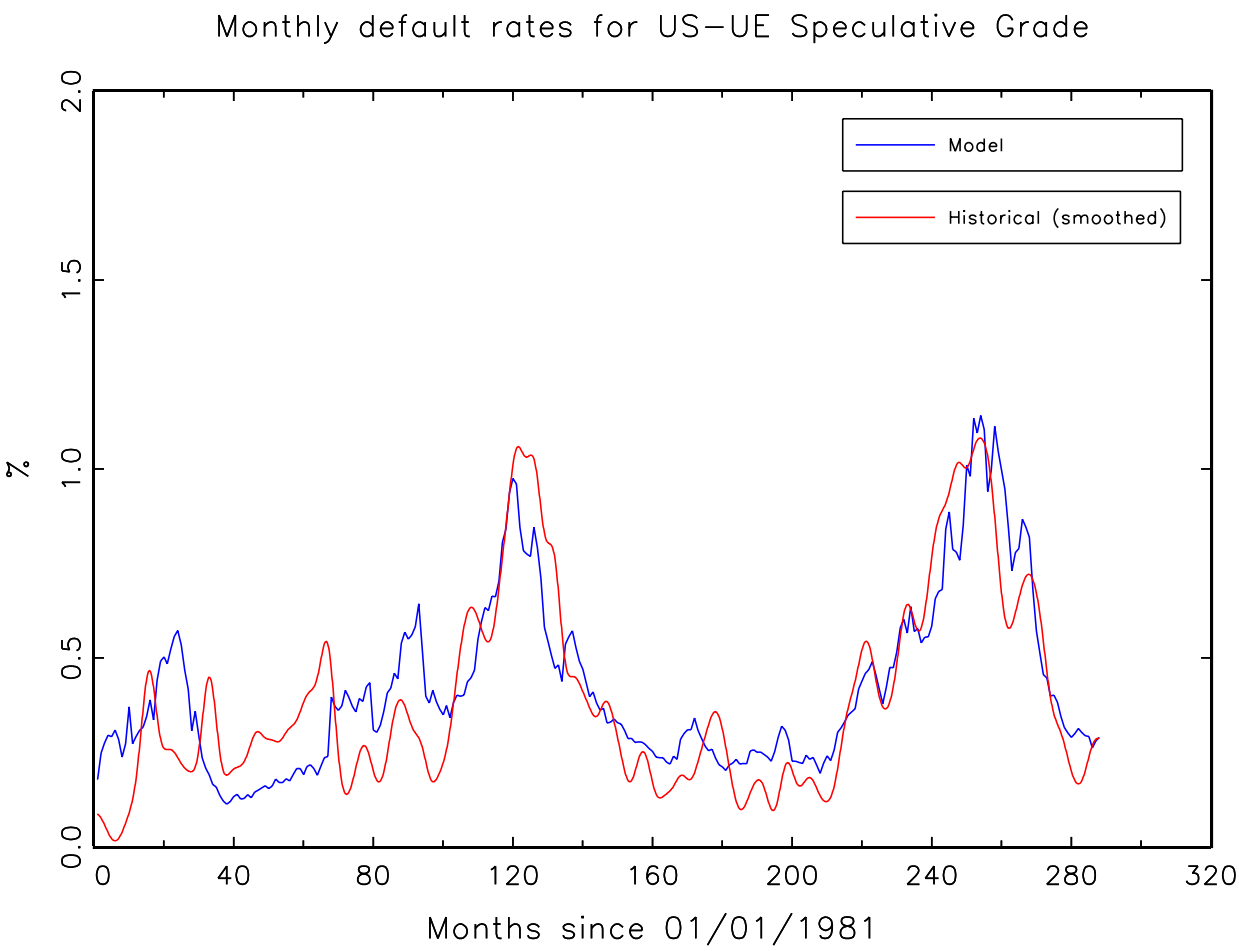

Figure 5.2: Monthly default rates, as provided by the basic model and empirically calculated

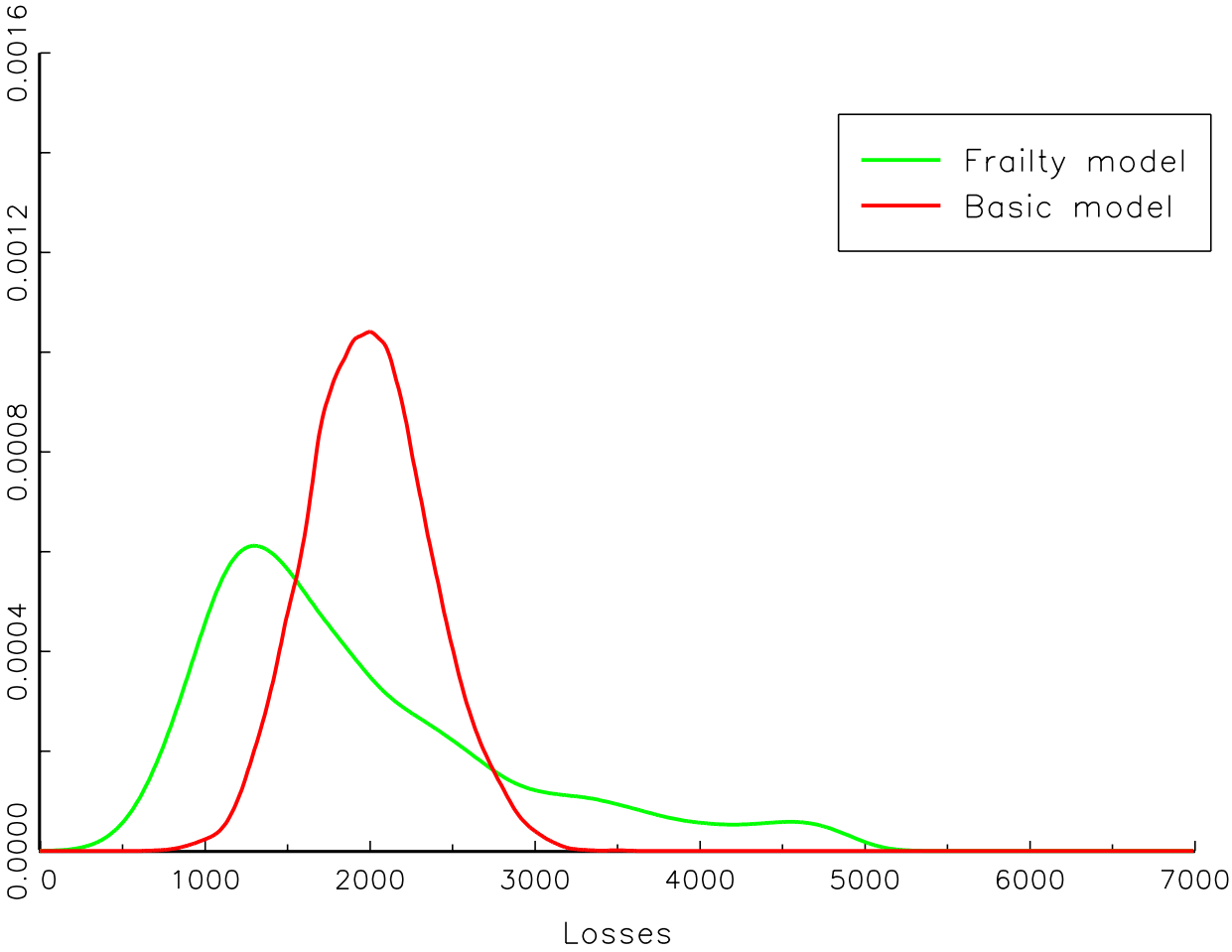

Figure 5.3: Loss distributions given by the basic model and the frailty model

Bibliographie

[1] A. Alfonsi. On the discretization schemes for the CIR (and Bessel squared) processes. *Monte Carlo Methods and Applications*, 11(4) :355–384, 2005.

[2] A. Alfonsi. High order discretization schemes for the CIR process : Application to affine term structure and Heston models. *Mathematics and Computations (to appear)*, 2009.

[3] L. Andersen. Efficient Simulation of the Heston Stochastic Volatility Model. *SSRN eLibrary*, 2007.

[4] P.K. Andersen, O. Borgan, R.D. Gill, and N. Keiding. *Statistical models based on counting processes*. Springer Series in Statistics. Springer-Verlag, New York, 1993.

[5] M. Avellaneda, D. Boyer-Olson, J. Busca, and P. Friz. Reconstructing volatility. *Risk*, pages 87–91, October 2002.

[6] G. Bakshi, N. Kapadia, and D. Madan. Stock return characteristics, skew laws, and the differential pricing of individual equity options. *Review of Financial Studies*, 16 :101–143, 2003.

[7] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function. *Probability Theory and Related Fields*, 104(1) :43–60, 1996.

[8] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. II. Convergence rate of the density. *Monte Carlo Methods and Applications*, 2(2) :93–128, 1996.

[9] A. Bangia, F.X. Diebold, and T. Schuermann. Ratings Migration and the Business Cycle, With Application to Credit Portfolio Stress Testing. Center for Financial Institutions Working Papers 00-26, Wharton School Center for Financial Institutions, University of Pennsylvania, 2000.

[10] H. Berestycki, J. Busca, and I. Florent. An inverse parabolic problem arising in finance. *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique*, 331(12) :965–969, 2000.

[11] H. Berestycki, J. Busca, and I. Florent. Asymptotics and calibration of local volatility models. *Quantitative Finance*, 2(1) :61–69, 2002.

[12] A. Berkaoui, M. Bossy, and A. Diop. Euler scheme for SDEs with non-Lipschitz diffusion coefficient : strong convergence. *ESAIM. Probability and Statistics*, 12 :1–11 (electronic), 2008.

[13] A. Beskos, O. Papaspiliopoulos, and G.O. Roberts. Retrospective exact simulation of diffusion sample paths. *Bernoulli*, 12(6), December 2006.

[14] A. Beskos, O. Papaspiliopoulos, G.O. Roberts, and P. Fearnhead. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 68(3) :333–382, 2006.

[15] A. Beskos and G.O. Roberts. Exact simulation of diffusions. *The Annals of Applied Probability*, 15(4) :2422–2444, 2005.

[16] N.P.B. Bollen and R.E. Whaley. Does net buying pressure affect the shape of implied volatility functions ? *Journal of Finance*, 59(2) :711–753, 04 2004.

[17] D. Bosq. *Nonparametric statistics for stochastic processes*, volume 110 of *Lecture Notes in Statistics*. Springer-Verlag, New York, second edition, 1998. Estimation and prediction.

[18] N. Branger and C. Schlag. Why is the index smile so steep ? *Review of Finance*, 8(1) :109–127, 2004.

[19] M. Broadie and P. Glasserman. Estimating security price derivatives using simulation. *Management Science*, 42(2) :269–285, 1996.

[20] M. Broadie and Ö. Kaya. Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research*, 54(2) :217–231, 2006.

[21] G. Celeux and J. Diebolt. The SEM algorithm : a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly*, 2 :73–82, 1985.

[22] P. Cizeau, M. Potters, and J-P. Bouchaud. Correlation structure of extreme stock returns. *Quantitative Finance*, 1(2) :217–222, February 2001.

[23] D. Clayton and J. Cuzick. Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society. Series A. General*, 148(2) :82–117, 1985.

[24] R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, and D.E. Knuth. On the lambert W function. *Advances in Computational Mathematics*, 5 :329–359, 1996.

[25] F. Couderc and O. Renault. Times-to-default :life cycle, global and industry cycle impact. FAME Research Paper Series rp142, International Center for Financial Asset Management and Engineering, March 2005.

[26] M. Crouhy, D. Galai, and R. Mark. A comparative analysis of current credit risk models. *Journal of Banking & Finance*, 24(1-2) :59–117, January 2000.

[27] A.B. Cruzeiro, P. Malliavin, and A. Thalmaier. Geometrization of Monte-Carlo numerical analysis of an elliptic operator : strong approximation. *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique*, 338(6) :481–486, 2004.

[28] A. De Servigny and O. Renault. Default correlation : empirical evidence. S&P working paper, 2002.

[29] G. Deelstra and F. Delbaen. Convergence of discretized stochastic (interest rate) processes with stochastic drift term. *Applied Stochastic Models and Data Analysis*, 14(1) :77–84, 1998.

[30] M. Delloye, J-D. Fermanian, and M. Sbai. Dynamic frailties and credit portfolio modelling. *Risk*, 19(10) :100–105, October 2006.

[31] A. Dermoune. Propagation and conditional propagation of chaos for pressureless gas equations. *Probability Theory and Related Fields*, 126(4) :459–476, 2003.

[32] J. Diebolt and E.H.S. Ip. Stochastic EM : method and application. In *Markov chain Monte Carlo in practice*, Interdiscip. Statist., pages 259–273. Chapman & Hall, London, 1996.

[33] H. Doss. Liens entre équations différentielles stochastiques et ordinaires. *Ann. Inst. H. Poincaré Sect. B (N.S.)*, 13(2) :99–125, 1977.

[34] F. Dubois and T. Lelievre. Efficient pricing of Asian options by the PDE approach. *Journal of Computational Finance*, 8(2), 2004.

[35] D. Duffie, A. Eckner, G. Horel, and L. Saita. Frailty correlated default. *The Journal of Finance (to appear)*, 2009.

[36] D. Duffie and K.J. Singleton. Modeling term structures of defaultable bonds. *Review of Financial Studies*, 12(4) :687–720, 1999.

[37] B. Dupire. Pricing with a smile. *Risk*, pages 18–20, January 1994.

[38] P. Fearnhead, O. Papaspiliopoulos, and G.O. Roberts. Particle filters for partially observed diffusions. *Working paper. Lancaster University.*, 2006.

[39] J-D. Fermanian and M. Sbai. A comparative analysis of dependence levels in intensity-based and Merton-style credit risk models. In Greg N. Gregoriou, editor, *Advances in Risk Management*. Palgrave MacMillan, 2006.

[40] R. Frey and A.J. McNeil. Modelling dependent defaults. Working paper, 2000.

[41] A. Friedman. *Partial differential equations of parabolic type*. Prentice-Hall Inc., Englewood Cliffs, N.J., 1964.

[42] M. Fu, D. Madan, and T. Wang. Pricing continuous Asian options : a comparison of Monte-Carlo and Laplace transform inversion methods. *Journal of Computational Finance*, 2(2), 1999.

[43] H. Geman, N. El Karoui, and J.C. Rochet. Changes of numéraires, changes of probability measure and option pricing. *Journal of Appliede Probability*, 32(2) :443–458, 1995.

[44] H. Geman and A. Eydeland. Domino effect. *Risk*, pages 65–67, April 1995.

[45] H. Geman and M. Yor. Bessel processes, Asian option and perpetuities. *Mathematical Finance*, 3(4), 1993.

[46] K. Giesecke and S. Azizpour. Self-exciting corporate defaults : Contagion vs. frailty. 2008.

[47] M. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In *Monte Carlo and quasi-Monte Carlo methods 2006*, pages 343–358. Springer, Berlin, 2008.

[48] M. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3) :607–617, 2008.

[49] E. Gobet. Weak approximation of killed diffusion using Euler schemes. *Stochastic Processes and their Applications*, 87(2) :167–197, 2000.

[50] E. Gobet and C. Labart. Sharp estimates for the convergence of the density of the Euler scheme in small time. *Electronic Communications in Probability*, 13 :352–363, 2008.

[51] M.B. Gordy. A comparative anatomy of credit risk models. *Journal of Banking & Finance*, 24(1-2) :119–149, 2000.

[52] J. Guyon. Euler scheme and tempered distributions. *Stochastic Processes and their Applications*, 116(6) :877–904, 2006.

[53] I. Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probability Theory and Related Fields*, 71(4) :501–516, 1986.

[54] J. Helwege and P. Kleiman. Understanding Aggregate Default Rates of High Yield Bonds. *Current Issues in Economics and Finance*, 2(6), May 1996.

[55] S.L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2) :327–43, 1993.

[56] P. Hougaard. *Analysis of multivariate survival data*. Statistics for Biology and Health. Springer-Verlag, New York, 2000.

[57] J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2) :281–300, 1987.

[58] J.E. Ingersoll. *Theory of Financial Decision Making*. Rowman & Littlefield, 1987.

[59] R.A. Jarrow, D. Lando, and S.M. Turnbull. A Markov model for the term structure of credit risk spreads. *The Review of Financial Studies*, 10(2) :481–523, 1997.

[60] B. Jourdain and M. Sbai. Exact retrospective Monte Carlo computation of arithmetic average asian options. *Monte Carlo Methods and Applications*, 13(2) :135–171, 2007.

[61] C. Kahl and P. Jäckel. Fast strong approximation Monte Carlo schemes for stochastic volatility models. *Quantitative Finance*, 6(6) :513–536, December 2006.

[62] C. Kahl and H. Schurz. Balanced Milstein methods for ordinary SDEs. *Monte Carlo Methods and Applications*, 12(2) :143–170, 2006.

[63] I. Karatzas and S.E. Shreve. *Brownian motion and stochastic calculus*. Springer-Verlag New-York, second edition, 1991.

[64] D. Kavvathas. Estimating Credit Rating Transition Probabilities for Corporate Bonds. Working paper series, University of Chicago, 2000.

[65] A. Kebaier. Statistical Romberg extrapolation : a new variance reduction method and applications to option pricing. *The Annals of Applied Probability*, 15(4) :2681–2705, 2005.

[66] S.C. Keenan, J.R. Sobehart, and D.T. Hamilton. Predicting Default Rates : A Forecasting Model for Moody's Issuer-Based Default Rates. Working paper, 1999.

[67] A. Kemna and A. Vorst. A pricing method for options based on average asset values. *Journal of Banking and Finance*, 14(1) :113–129, 1990.

[68] J. Kim. A way to condition transition matrix on wind. Working paper, May 1999.

[69] A. Kohatsu-Higa. Weak approximations. A Malliavin calculus approach. *Mathematics of Computation*, 70(233) :135–172, 2001.

[70] V. Konakov and E. Mammen. Edgeworth type expansions for Euler schemes for stochastic differential equations. *Monte Carlo Methods and Applications*, 8(3) :271–285, 2002.

[71] S.J. Koopman, A. Lucas, and A. Monteiro. The multi-state latent factor intensity model for credit rating transitions. Working paper, Tinbergen Institute, Amsterdam, 2005.

[72] H.U. Koyluoglu and A. Hickman. A generalized framework for credit risk portfolio models. Working paper, 1998.

[73] H.U. Koyluoglu and A. Hickman. Reconcilable differences. *Risk*, 11(10) :56–62, October 1998.

[74] H. Kunita. Stochastic differential equations and stochastic flows of diffeomorphisms. In *École d'été de probabilités de Saint-Flour, XII—1982*, volume 1097 of *Lecture Notes in Math.*, pages 143–303. Springer, Berlin, 1984.

[75] T.G. Kurtz and P. Protter. Wong-Zakai corrections, random evolutions, and simulation schemes for SDEs. In *Stochastic analysis*, pages 331–346. Academic Press, Boston, MA, 1991.

[76] S. Kusuoka. Approximation of expectation of diffusion process and mathematical finance. In *Taniguchi Conference on Mathematics Nara '98*, volume 31 of *Adv. Stud. Pure Math.*, pages 147–165. Math. Soc. Japan, Tokyo, 2001.

[77] S. Kusuoka. Approximation of expectation of diffusion processes based on Lie algebra and Malliavin calculus. In *Advances in mathematical economics. Vol. 6*, volume 6 of *Adv. Math. Econ.*, pages 69–83. Springer, Tokyo, 2004.

[78] S. Kusuoka and D. Stroock. Applications of the Malliavin calculus. I. In *Stochastic analysis (Katata/Kyoto, 1982)*, volume 32 of *North-Holland Math. Library*, pages 271–306. North-Holland, Amsterdam, 1984.

[79] D. Lando. On cox processes and credit risky securities. *Review of Derivatives Research*, 2 :99–120, 1998.

[80] D. Lando and T.M. Skødeberg. Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking & Finance*, 26(2-3) :423–444, 2002.

[81] B. Lapeyre and E. Temam. Competitive Monte Carlo methods for pricing Asian options. *Journal of Computational Finance*, 5(1), 2001.

[82] P. Lee, L. Wang, and A. Kerim. Index volatility surface via moment-matching techniques. *Risk*, pages 85–89, December 2003.

[83] E. Levy. Pricing european average rate currency options. *Journal of International Money and Finance*, 11(5) :474–491, October 1992.

[84] F.A. Longstaff and E.S. Schwartz. Valuing American options by simulation : a simple least-squares approach. *Review of Financial Studies*, 14(1) :113–147, 2001.

[85] R. Lord. Partially exact and bounded approximations for arithmetic Asian options. *Journal of Computational Finance*, 10(2), 2006.

[86] R. Lord, R. Koekkoek, and D.J. Van Dijk. A Comparison of Biased Simulation Schemes for Stochastic Volatility Models. *SSRN eLibrary*, 2008.

[87] T. Lyons and N. Victoir. Cubature on Wiener space. *Proceedings of The Royal Society of London. Series A. Mathematical, Physical and Engineering Sciences*, 460(2041) :169–198, 2004. Stochastic analysis with applications to mathematical finance.

[88] G.J. McLachlan and T. Krishnan. *The EM algorithm and extensions*. Wiley Series in Probability and Statistics : Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1997. A Wiley-Interscience Publication.

[89] S. Méléard. Asymptotic behaviour of some interacting particle systems ; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995)*, volume 1627 of *Lecture Notes in Math.*, pages 42–95. Springer, Berlin, 1996.

[90] R.C. Merton. On the pricing of corporate debt : The risk structure of interest rates. *The Journal of Finance*, 29(2) :449–470, 1974.

[91] B. Metayer. Semi-parametric cox type regression model for credit rating transition probabilities estimation. Working paper, 2004.

[92] G.N. Milstein. *Numerical Integration of Stochastic Differential Equations*, volume 313 of *Mathematics and its Applications*. Kluwer Academic Publishers, 1995.

[93] E.A. Nadaraya. On estimating regression. *Theory of Probability and its Applications*, 9(1) :141–142, 1964.

[94] P. Nickell, W. Perraudin, and S. Varotto. Stability of ratings transitions. *Bank of England Quarterly Bulletin*, 41 :216–216, 2001.

[95] S. Ninomiya. A new simulation scheme of diffusion processes : application of the Kusuoka approximation to finance problems. *Mathematics and Computers in Simulation*, 62(3-6) :479–486, 2003.

[96] S. Ninomiya. A partial sampling method applied to the Kusuoka approximation. *Monte Carlo Methods and Applications*, 9(1) :27–38, 2003.

[97] S. Ninomiya and M. Ninomiya. A new higher-order weak approximation scheme for stochastic differential equations and the Runge-Kutta method. *Finance and Stochastics*, 13 :415–443, 2009.

[98] S. Ninomiya and N. Victoir. Weak approximation of stochastic differential equations and application to derivative pricing. *Applied Mathematical Finance*, 15(1-2) :107–121, 2008.

[99] Basel Committee on Banking Supervision. International convergence of capital measurement and capital standards. Report, June 2004.

[100] M.C. Paik, W. Tsai, and R. Ottman. Multivariate survival analysis using piecewise gamma frailty. *Biometrics*, 50(4) :975–988, 1994.

[101] S.T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, 1998. Theory.

[102] S.T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. II.* Probability and its Applications (New York). Springer-Verlag, New York, 1998. Applications.

[103] D. Revuz and M. Yor. *Continuous martingales and Brownian motion.* Springer-Verlag Berlin Heidelberg, 1991.

[104] C. P. Robert and G. Casella. *Monte Carlo statistical methods.* Springer Texts in Statistics. Springer-Verlag, New York, 1999.

[105] L.C.G. Rogers and Z. Shi. The value of an Asian option. *Jounal of Applied Probability*, 32(4) :1077–1088, 1995.

[106] M. Romano and N. Touzi. Contingent claims and market completeness in a stochastic volatility model. *Mathematical Finance*, 7(4) :399–412, 1997.

[107] W.J. Runggaldier and C. Fontana. Credit risk and incomplete information : linear filtering and em parameter estimation. 2009.

[108] W.J. Runggaldier and R. Frey. Credit risk and incomplete information : a nonlinear-filtering approach. *Finance and Stochastics (to appear)*, 2009.

[109] T.H. Rydberg. A note on the existence of unique equivalent martingale measures in a markovian setting. *Finance and Stochastics*, 1(3) :251–257, 1997.

[110] P. Schönbucher. Factor models for credit portfolio credit risk. Working paper, 2000.

[111] P. Schönbucher. Information-driven default contagion. Working paper, 2003.

[112] L.O. Scott. Option pricing when the variance changes randomly : theory, estimation, and an application. *The Journal of Financial and Quantitative Analysis*, 22(4) :419–438, 1987.

[113] P. Seumen Tonou. *Méthodes numériques probabilistes pour la résolution d'équations du transport et pour l'évaluation d'options exotiques.* Phd thesis, Université de Provence, Aix-Marseille 1, 1997.

[114] W.F. Sharpe. Capital asset prices : A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3) :425–442, 1964.

[115] E.M. Stein and J.C. Stein. Stock price distributions with stochastic volatility : an analytic approach. *Review of Financial Studies*, 4(4) :727–752, 1991.

[116] A-S. Sznitman. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Math.*, pages 165–251. Springer, Berlin, 1991.

[117] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Analysis and Applications*, 8(4) :483–509, 1990.

[118] D. Talay and O. Vaillant. A stochastic particle method with random weights for the computation of statistical solutions of McKean-Vlasov equations. *The Annals of Applied Probability*, 13(1) :140–180, 2003.

[119] H. Tanaka and A. Kohatsu-Higa. An operator approach for Markov chain weak approximations with an application to infinite activity Lévy driven SDEs. *The Annals of Applied Probability (to appear)*, 2009.

[120] E. Tanré. *Étude probabiliste des équations de Smoluchowski ; Schéma d'Euler pour des fonctionnelles ; Amplitude du mouvement brownien avec dérive.* Phd thesis, Université Henri Poincaré Nancy 1, 2001.

[121] E. Temam. *Couverture approchée d'options exotiques - Pricing des options asiatiques.* Phd thesis, Université Paris 6, 2001.

[122] S. Turnball and L. Wakeman. A quick algorithm for pricing European average options. *Journal of Financial and Quantitative Analysis*, 16 :377–389, 1991.

[123] J. Vecer. A new PDE approach for pricing arithmetic Asian options. *Journal of Computational Finance*, 4(4), 2001.

[124] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics.* American Mathematical Society, 2003.

[125] T. Vorst. Prices and hedge ratios of average exchange rate options. *International Review of Financial Analysis*, 1(3) :179–193, 1992.

[126] W. Wagner. Unbiased Monte Carlo evaluation of certain functional integrals. *Journal of Computational Physics*, 71(1) :21–33, 1987.

[127] W. Wagner. Monte Carlo evaluation of functionals of solutions of stochastic differential equations. Variance reduction and numerical examples. *Stochastic Analysis and Applications*, 6(4) :447–468, 1988.

[128] W. Wagner. Unbiased multi-step estimators for the Monte Carlo evaluation of certain functional integrals. *Journal of Computational Physics*, 79(2) :336–352, 1988.

[129] W. Wagner. Unbiased Monte Carlo estimators for functionals of weak solutions of stochastic differential equations. *Stochastics Stochastics Reports*, 28(1) :1–20, 1989.

[130] G.S. Watson. Smooth regression analysis. *Sankhyā (Statistics). The Indian Journal of Statistics. Series A*, 26 :359–372, 1964.

[131] J.B. Wiggins. Option values under stochastic volatility : Theory and empirical estimates. *Journal of Financial Economics*, 19(2) :351–372, 1987.

[132] T.C. Wilson. Portfolio credit risk i. *Risk*, pages 111–117, September 1997.

[133] T.C. Wilson. Portfolio credit risk ii. *Risk*, pages 56–61, October 1997.

[134] B. Wong and C.C. Heyde. On the martingale property of stochastic exponentials. *Journal of Applied Probability*, 41(3) :654–664, 2004.

[135] K.K.W. Yau and C.A. McGilchrist. Ml and reml estimation in survival analysis with time dependent correlated frailty. *Statistics in Medicine*, 17(11) :1201–1213, 1998.

[136] F. Yu. Correlated defaults in intensity-based models. Working paper, University of California, Irvine, 2002.

[137] F. Yu. Default correlation in reduced-form models. Working paper, University of California, Irvine, 2003.

[138] H. Yue and K.S. Chan. A dynamic frailty model for multivariate survival data. *Biometrics*, 53(3) :785–793, 1997.

Ce document a été préparé à l'aide de LaTeX 2_ε.